OM protein - protein search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: March 12, 2003, 00:08:56 ; Search time 34 Seconds
(without alignments)
2830.310 Million cell updates/sec

Post-processing:

Title: US-10-046-433-40
Perfect score: 5506
Sequence: 1 MAEPGHSHHILSARVRGRTER..........LGRSNILPPRGLLMDLTQCR 1001

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB	ID	Description
1	251	4.6	1609	1	MMHUB2	laminin gamma-1 ch
2	225	4.1	1607	1	MMMSB2	laminin gamma-1 ch
3	224.5	4.1	1639	1	MMFFB2	laminin gamma-1 ch
4	211.5	3.8	713	1	A35502	major surface-labe
5	209	3.8	1548	1	S34583	serine proteinase
6	197.5	3.6	1111	2	T26972	hypothetical prote
7	192	3.5	1557	2	T28811	hypothetical prote
8	189	3.4	677	2	C42125	trophozoite cystei
9	183.5	3.4	1391	2	T20406	trophozoite cystei
10	179.5	3.3	1797	2	T21889	hypothetical prote
11	179.5	3.3	1805	2	T21888	hypothetical prote
12	178	3.2	1274	2	T42017	cysteine rich prot
13	177.5	3.2	1766	2	A42125	trophozoite cystei
14	177	3.2	557	2	A48434	variant-specific s
15	175	3.2	2352	2	T30201	Notch homolog prot
16	174	3.2	1786	1	MMHUB1	laminin beta-1 cha
17	172.5	3.1	3106	1	S53868	laminin alpha-2 ch
18	172	3.1	1372	2	T25933	hypothetical prote
19	171.5	3.1	667	2	A48579	trophozoite surfac
20	170.5	3.1	1786	1	MMMSB1	laminin beta-1 cha
21	170	3.1	2823	2	T23064	hypothetical prote
22	170	3.1	2823	2	F87908	protein T22A3.8 [i
23	170	3.1	3102	2	T43291	laminin alpha chai
24	169	3.1	1680	1	A43434	laminin alpha-2 ch
25	167	3.0	610	2	A35046	furin (EC 3.4.21.7
26	166.5	3.0	1574	2	T13954	E-selectin precurs
27	163.5	3.0	1790	1	MMFFB1	laminin beta-1 cha
28	163.5	3.0	2090	2	T30075	hypothetical prote
29	163	3.0	2165	2	T21371	hypothetical prote
30	163	3.0	3084	1	MMMSA	laminin alpha-1 ch
31	161.5	2.9	2524	2	A35844	Xotch protein - Af
32	161	2.9	4391	2	A38096	perlecan precursor
33	160.5	2.9	1193	2	A44018	laminin B2t chain
34	160.5	2.9	3672	2	T23433	hypothetical prote
35	160.5	2.9	3704	2	T37316	probable laminin a
36	160	2.9	993	2	I48653	mouse developmenta
37	159	2.9	1169	2	S38181	flocculation prote
38	159	2.9	2761	2	T21064	hypothetical prote
39	158	2.9	1203	2	A49175	Motch B protein -
40	158	2.9	2871	2	A55567	fibrillin I - bovi
41	158	2.9	3712	2	S18253	fibrillin-1 ch
42	157.5	2.9	2907	2	A57278	fibrillin-2 precur
43	157.5	2.9	2918	2	A54105	fibrillin-2 precur
44	157.5	2.9	3075	2	S14458	laminin alpha-1 ch
45	156	2.8	1220	2	A56136	jagged protein pre

ALIGNMENTS

RESULT 1
MMHUB2
laminin gamma-1 chain precursor - human
N;Alternate names: laminin chain B2
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
C;Accession: S13548; A28158; S13549; B34961; S14664; S23567
R;Kallunki, T.; Ikonen, J.; Chow, L.T.; Kallunki, P.; Tryggvason, K.
J. Biol. Chem. 266, 221-228, 1991
A;Title: Structure of the human laminin B2 gene reveals extensive divergence fr
A;Reference number: S13548; MUID:91093128; PMID:1985895
A;Accession: S13548
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1609 <KAL>
A;Cross-references: GB:M55217; NID:g186937
A;Note: the nucleotide sequence was submitted to GenBank, February 1991
R;Pikkarainen, T.; Kallunki, T.; Tryggvason, K.
J. Biol. Chem. 263, 6751-6758, 1988
A;Title: Human laminin B2 chain. Comparison of the complete amino acid sequence with
A;Reference number: A28158; MUID:88198245; PMID:3360804
A;Molecule type: mRNA
A;Residues: 1-211, I',213-1609 <PIK>
A;Cross-references: EMBL:J03202; NID:g186916; PIDN:AAA59488.1; PID:g307107
R;Olsen, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasak
Lab. Invest. 60, 772-782, 1989
A;Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2
A;Reference number: A34961; MUID:89280632; PMID:2733383
A;Molecule type: mRNA
A;Residues: 868-1551,'N',1553-1609 <OLS>
A;Cross-references: EMBL:X13939; NID:g34237; PIDN:CAA32122.1; PID:g34238
R;Santos, C.L.S.; Sabbaga, J.; Brentani, R.
DNA Seq. 1, 275-277, 1991
A;Title: Differences in human laminin B2 sequences.
A;Reference number: S14664; MUID:92216129; PMID:1806043
A;Accession: S14664
A;Molecule type: mRNA
A;Residues: 1282-1609 <SAN>
A;Cross-references: EMBL:X13939; NID:g34237; PIDN:CAA32122.1; PID:g34238
R;Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Pikkarainen, T.; Tryggvason, K
A;Title: Genes for the human laminin B1 and B2 chains.
A;Reference number: S23566

```
A:Accession: S23567
A:Molecule type: DNA
A:Residues: 801-1481,'R',1483-1609 <VUO>
A:Note: mRNA was also sequenced
A:Genetics:
A:Gene: GDB:LAMC1; LAMB2
A:Cross-references: GDB:120136; OMIM:150290
A:Map position: 1q31-1q31
A:Introns: 140/1; 241/3; 285/2; 341/1; 404/1; 443/2; 476/2; 522/1; 563/1; 626/2; 664/1;
/3; 1525/1
A:34-285/Domain: VI <DOM6>
A:1-33/Product: laminin gamma-1 chain #status predicted <MAT>
A:1-33/Domain: signal sequence #status predicted <SIG>
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Description: interact with cells and with other basement membrane proteins to promote
C:Function:
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
F;34-285/Domain: VI <DOM6>
F;286-504/Domain: V <DOM5>
F;286-339/Domain: laminin-type EGF-like homology <LE01>
F;342-395/Domain: laminin-type EGF-like homology <LE02>
F;398-442/Domain: laminin-type EGF-like homology <LE03>
F;445-492/Domain: laminin-type EGF-like homology <LE04>
F;495-504/Domain: laminin-type EGF-like homology #status atypical <LE05>
F;505-689/Domain: IV <DOM4>
F;690-1034/Domain: III <DOM3>
F;690-721/Domain: laminin-type EGF-like homology #status atypical <LE06>
F;724-770/Domain: laminin-type EGF-like homology <LE07>
F;773-825/Domain: laminin-type EGF-like homology <LE08>
F;828-881/Domain: laminin-type EGF-like homology <LE09>
F;884-932/Domain: laminin-type EGF-like homology <LE10>
F;935-980/Domain: laminin-type EGF-like homology <LE11>
F;983-1028/Domain: laminin-type EGF-like homology <LE12>
F;1035-1609/Domain: II/I <DOM1>
F;1035-1609/Region: heptad repeats
F;40-50/Disulfide bonds: #status predicted
F;60,134,576,650,1022,1107,1161,1175,1205,1223,1241,1380,1395,1439/Binding site: carbohy
F;1031,1034,1600/Disulfide bonds: interchain #status predicted
```

```
Query Match      4.6%;  Score 251;  DB 1;  Length 1609;
Best Local Similarity   19.6%;   Pred. No. 4.4e-08;
Matches  212;  Conservative  85;  Mismatches  382;  Indels 400;  Gaps  54;

Qy     23 PRLWRLLIWAGTAFQVT------------QGTGPELHACKES------------EYHYEYTACDSTGSSRW 68
          :|| ::||                     |||||  :|||:            ::| |||
Db    130 PSSINLTLHLGKAFDITYVRLKFHTSRPESPAIYKRTREDGPWIPYQYYSSCENYYSKA 189

Qy     69 RVAVPHTPG-----LCTSLPDPVKGTECSFSCNAGEFLDMKDQSCKPCAEGYRYSLGTGIR 123
          | | |        ||| | |  |   |:|:|||  ||  || :: |
Db    190 NRGFIRTGGDEQQALCTD-----EFSDF-----SPLTGGNVAFST--- 224

Qy    124 FDEWDELPHGFASLSANMELDDSAAESTGNCTSSKKWVPRGDYIAFNTDECTATLMYAVNL 183
          |   |:|  |   | |||:|| ||    |::|            | |  :|
Db    225 -----------LEGRPSAYNFEDNSPVLQEWVTATD----- 253

Qy    184 KQSGTVNFE----------YYPDSSTIFEEFVQNDQCQPNADDSRMKTTEKGWEEHSV 233
          |  ||| |:         :: ||:: ||    ||  |:    | |:|  |:|
Db    254 NRLNTFGDEVFNDPKYLKSIYYAISD-----FAVGRCKCNGHASECMKN---EFDKL 303

Qy    234 ELNRGNNVL-----------YWRTTAEFSVWTKVPKPVLVRNIAITGVAYTSECFPC 278
          | | |::|           :|:|||  | ||| |   ||  :|    |  |:|  |
Db    304 VCNCKHNTYGVDCEKCLPFNDRPWRRATAES-----ASECLPC 342

Qy    279 -------KPGTV-----ADKQGSSPCKLCPANSY---SNKGETSCHQCD 312
          |: ||  |  :|  |:| |  :||| :|      |||| |:| :
Db    343 DCNGRSQECYFDPELYRSTGHGGHCTNCQDNTDGAHCERCRENFFRLGNNEACSSCH-CS 401

Qy    313 P--------DKYSEKGSSSSCNVRPA-----CTDKDYFYTHT----ACDANGETQLM 351
          |        ||: |:    :|| ||      ||:||| |::    ||| |||  |:
Db    402 PVGSLSTQCDSYGR-----CSCKPGYWGDKDCDRCQFGFHSLTEAGCRPCSCDPSGSIDEC 456

Qy    352 YKWAKPKICSEDLEGAVKLPASGVKTHCPPCNPGFFKTNNST----CQPCPYGYSYNGSSDC 408
          :| : :|| |:|   | ::|||||:|||  ||||||||| :|    |:||| |||| ||| |
```

```
RESULT 2
MMSB2
laminin gamma-1 chain precursor - mouse
N:Alternate names: laminin chain B2
C:Species: Mus musculus (house mouse)
C:Date: 28-Feb-1986 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
C:Accession: A28469; A27729; A28082; S02680; S05327; A02870; S13544; S14552
R:Sasaki, M.; Yamada, Y.
J. Biol. Chem. 262, 17111-17117, 1987
A:Title: The laminin B2 chain has a multidomain structure homologous to the B1 chain.
A:Reference number: A28469; MUID:88059118; PMID:3680290
A:Molecule type: mRNA
A:Residues: 1-1607 <SAS>
R:Durkin, M.E.; Bartos, B.B.; Liu, S.H.; Phillips, S.L.; Chung, A.E.
Biochemistry 27, 5198-5204, 1988
A:Cross-references: EMBL:J03484; NID:g1986694; PIDN:AAA39405.1; PID:g293688
R:Durkin, M.E.; Bartos, B.B.; Liu, S.H.; Phillips, S.L.; Chung, A.E.
Biochemistry 27, 5198-5204, 1988
A:Title: Primary structure of the mouse laminin B2 chain and comparison with laminin
A:Cross-references: EMBL:J02930; NID:g198702; PIDN:AAA39408.1; PID:g293691
A:Reference number: A27729; MUID:89000737; PMID:3167041
A:Accession: A27729
A:Molecule type: mRNA
R:Ogawa, K.; Burbelo, P.D.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 263, 8384-8389, 1988
A:Cross-references: EMBL:J03484,'D',265-336,'C',338-446,'PS',449-661,'S',663-885,887-1155,1157-1433
A:Residues: 1-263,'D',265-336,'C',338-446,'PS',449-661,'S',663-885,887-1155,1157-1433
A:Title: The laminin B2 chain promoter contains unique repeat sequences and is active
A:Note: the authors translated the codon TAT for residue 544 as Asp and GCG for resid
A:Reference number: A28082; MUID:88228071; PMID:2836421
A:Accession: A28082
A:Molecule type: DNA
```

```
Db    457 NVETGRCVCKDNVEG----------FNCERCKPGFFNLESSNPRGCTPC--FCFGHSSVC 504
          |:  ||| | :||:           |||||||||| | ||::|| |:||  ||||||| |
Qy    409 TRCPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEYKGMTGWEVAGDHIYTAAGASDND 468

Db    505 TN----AVGISV-----YSISSTFQIDEDGWRAEQRDGSEASLEWSSEROD 546

Qy    469 FMILT------LVVPGFRPPQSVMADYFENKEVARITFVFET--LCSVNCELYFMVGVN 518
          ||:|:      |:||||| || |||||||:|||||||  ::   |||:|||:|| | ||
Db    547 IAVISDSYFPRYFIAPAKPFLGKOVLSYGGN-----LSFSFRVDRDTRLSAEDLVLEGAG 601

Qy    519 SRINTPVETWKGS-KGKQSYTYIIEENTTISFTW-----AFQ-----RTTFH 559
          ||:||| |||||| ||||||||:||||||:||||      ||      |||||
Db    602 LRVSVPLIAQGNSYPSETTVKVYVFERLHEATDYPWRPALTPFEFQKLLNNLTSIKIRGTYS 661

Qy    560 EASRKYTNDVAKI-----------YSIN 576
          | |  | |:| |:           |||
Db    662 ERSAGYLDDVTLASARPGPGVPATWVESCTCPVGYGGQFCEMCLSGYRRETPNLGPYSPC 721

Qy    577 VTNVMVNGVASYCRP-----CALEASDVGSSCTSCPAGYYIDRDSGT--CHSCP-PNTILK 628
          |:|  :||:||||      | ||:|| |||||:||||||||:|||:   ||:|| ||:||
Db    722 VLCACNGHSETCDPETGVCNCRDNTAGPHCEKCSDGYYGDSTAGTSSDCQPCPCPGGSSC 781

Qy    629 AHQPYGVQ-ACVPCGPGTKNNKIHSLC---YNDCTFSRNTPTR-----TFNYNFSALA 677
          |  ||||  ||  |||||:|:|||||:   ||||||||||||:     |||| || ||
Db    782 AVVPKTKEVVCTNCPTGYTGKRC-ELCDDGYFGDPLGRNGPVRLCRLCQCSDNIDPNAVG 840

Qy    678 NTVTLAGGPSFTSKGLKYFHH-----FTLSLCGNQGRKMSVCTDNVTDLRIP 724
          |:  | ||  | :|| | :     :|:|:|||:  | | ||||| || ||
Db    841 NCNRLTG----ECLKCIYNTAGFYCDRCKDGFFGNPLAPNPADKCRACNCN-----P 888

Qy    725 EGESGFSKSITAYVCQAVIIPPEVTGYKAGV----SSQPVSLAD-RLIGVTTDM- 773
          :| |  :|:||| ||||:|||||||||||     |||||:||| ||||||||:
Db    889 YGTMKQQSSCNPVTGQCECL-PHVTGQDCGACDPGFYNLQSGQGCERCDCHALGSTNGQC 947

Qy    774 -------TLDGIITSPAELFHLESLGIPDVIFYRSNDVTQSCSSGRSTTIRVRCSPQKT 825
          ||| ||||||| ||:|:|:|||:|:||:| |:|||| ||:|  |||:|: | ||| ||
Db    948 DIRTGQCECQPGITGQ----HCERC---EVNHFGFGPEGCKPCD----CHPE--- 988

Qy    826 VPGSLLL------PGTCS-----DGTCDGCNFHFL---WESAAACPLC------SVADY 864
          ||||:|       ||||     ||||| |:|  |    |||||||||       || |:
Db    989 --GSLSLQCKDDGRCECREGFVGNRCDQCEENFYNRSWPGCQQCPACYRLVKDKVADH 1045
```

A;Residues: 1-215,'A',217-239 <OGA>
A;Cross-references: EMBL:J03749; NID:g198704; PIDN:AAA39409.1; PID:g554184
A;Accession: S02680
A;Reference number: S02678; MUID:88326259; PMID:2458101
A;Title: Structure and distribution of N-linked oligosaccharide chains on various domain
A;Biochem. J. 252, 453-461, 1988
R;Fujiwara, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.
A;Cross-references: EMBL:J03749; NID:g198704; PIDN:AAA39409.1; PID:g554184
A;Molecule type: protein
A;Residues: 227-238 <FUJ>
R;Hartl, L.; Oberbaeumer, I.; Deutzmann, R.
Eur. J. Biochem. 173, 629-635, 1988
R;Barlow, D.P.; Green, N.M.; Kurkinen, M.; Hogan, B.L.M.
A;Title: The N terminus of laminin A chain is homologous to the B chains.
A;Reference number: S00624; MUID:88225080; PMID:3267223
A;Accession: S05327
A;Molecule type: protein
A;Residues: 227-238; 387-393,'F',395-405,881-912;1022-1034 <HAR>
R;Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbaeumer, I.; Hartl, L.
Eur. J. Biochem. 177, 35-45, 1988
A;Title: Structural study of long arm fragments of laminin. Evidence for repetitive C-te
A;Reference number: S01790; MUID:89030693; PMID:3181157
A;Accession: S02037
A;Molecule type: protein
A;Residues: 1362-1377,'X',1379-1392,'X',1394-1406 <DEU>
EMBO J. 3, 2355-2362, 1984
R;Olsen, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasaki,
Lab. Invest. 60, 772-782, 1989
A;Title: Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil a
R;Paulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.
EMBO J. 4, 309-316, 1985
A;Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2 cha
A;Reference number: A34961; MUID:89280632; PMID:2733383
A;Accession: S14552
A;Molecule type: protein
A;Residues: 881-912;1022-1034;1364-1377;1379-1392;1394-1409;1506-1525;1593-1606 <OLS>
A;Molecule type: protein
A;Residues: 1506-1523,'X',1525 <PAU>
R;Olsen, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasaki,
A;Accession: S13544
A;Reference number: S13543; MUID:85257455; PMID:3848400
A;Title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin.
A;Molecule type: mRNA
A;Accession: A02870
A;Reference number: A02870; MUID:85051302; PMID:6209134
A;Title: Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil
R;Barlow, D.P.; Green, N.M.; Kurkinen, M.; Hogan, B.L.M.
EMBO J. 3, 2355-2362, 1984
A;Molecule type: protein
A;Residues: 1391-1474,'K',1476-1575,'N',1577-1607 <BAR>
A;Accession: A02870
A;Molecule type: protein
A;Residues: 1362-1377,'X',1379-1392,'X',1394-1406 <DEU>
A;Accession: S02037
A;Reference number: S01790; MUID:89030693; PMID:3181157
A;Title: Structural study of long arm fragments of laminin. Evidence for repetitive C-te
Eur. J. Biochem. 177, 35-45, 1988
R;Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbaeumer, I.; Hartl, L.
A;Residues: 227-238; 387-393,'F',395-405,881-912;1022-1034 <HAR>
A;Molecule type: protein
A;Accession: S05327
A;Reference number: S00624; MUID:88225080; PMID:3267223
A;Title: The N terminus of laminin A chain is homologous to the B chains.
A;Cross-references: EMBL:X05211; NID:g52862; PIDN:CAA28838.1; PID:g817975
A;Accession: S13544
A;Molecule type: protein
A;Residues: 1506-1523,'X',1525 <PAU>
A;Map position: 1
A;Introns: 138/1; 239/3
A;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
C;Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C;Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
C;Description: interact with cells and with other basement membrane proteins to promote
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-1607/Product: laminin gamma-1 chain #status predicted <MAT>
F;34-283/Domain: VI <DOM6>
F;284-337/Domain: laminin-type EGF-like homology #status atypical <LE01>
F;284-337/Domain: V <DOM5>
F;340-393/Domain: laminin-type EGF-like homology <LE02>
F;396-440/Domain: laminin-type EGF-like homology <LE03>
F;443-490/Domain: laminin-type EGF-like homology <LE04>
F;493-502/Domain: laminin-type EGF-like homology #status atypical <LE05>
F;503-687/Domain: IV <DOM4>
F;503-687/Domain: laminin-type EGF-like homology #status atypical <LE05>
F;688-1032/Domain: III <DOM3>
F;688-719/Domain: laminin-type EGF-like homology #status atypical <LE06>
F;722-768/Domain: laminin-type EGF-like homology <LE07>
F;771-823/Domain: laminin-type EGF-like homology <LE08>
F;826-879/Domain: laminin-type EGF-like homology <LE09>
F;882-930/Domain: laminin-type EGF-like homology <LE10>
F;933-978/Domain: laminin-type EGF-like homology <LE11>
F;981-1026/Domain: laminin-type EGF-like homology <LE12>
F;1033-1607/Domain: II/I <DOM2>

A;Gene: Lamb-2
A;Genetics:
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
A;Molecule type: protein
A;Residues: 124-223 <DEU>

F;1033-1607/Region: heptad repeats
F;38-48/Disulfide bonds: #status predicted
F;58,132,574,648,1020,1105,1159,1173,1203,1221,1239,1437/Binding site: carbohydrate (
F;1029,1032/Disulfide bonds: interchain #status predicted
F;1029,1032/Disulfide bonds: interchain #status predicted
F;1378,1393/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;1598/Disulfide bonds: interchain (to chain B1) #status experimental

Query Match   4.1%;   Score 225;   DB 1;   Length 1607;
Best Local Similarity 19.6%;   Pred. No. 2.1e-06;
Matches 212;   Conservative 83;   Mismatches 368;   Indels 416;   Gaps 53;

Qy   23 PRLWRLLLWAGTPAFQVT-------QGTGPELHACKES----------EYHYEYTACDSTGSRW 68
          ::| | |: | | | || |:|| ||||:| |: | | | |||
Db   128 PNSINLTLHLGKAFDITYVRLKFHTSRPESFAIYKRTEDGPWIPYQYYSGSCENTYSKA 187

Qy   69 RVAVPHYPG------LCTSLPDPVKGTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIR 123
          | | | | || ||| || | |||| | | || |
Db   188 NRGFIRTGGDEQQALCTD---------------EFSDI----SPLTGGNVAFST--- 222

Qy   124 FDEWDELPHGFASLSANMELDDSAAESTGNCTSSKWVPRGDYIAFNTDECTATLMYAVNL 183
          : | | | | : || | |
Db   124 ----------------LEGRPSAYNFDNSPVLQEWVTATD-----------IRVTL 251

Qy   184 KQSGTVNFE----------YYPDSSIFEFYVQNDQCQPNADDSRWMKTTEKGWEFHSV 233
          | | | | | | | || | | | | | || |
Db   223 ---------------YYPDSSIFEFYVQNDQCQPNADDSRWMKTTEKGWEFHSV 233

Qy   234 ELNRGNNVL------------YWRTTAFSVWTKVF-------------KPVLVRNIAI 266
          | | | | | | | | | |||| | |
Db   252 NRLNTFGDEVFNEPKVLKSYYAISD-----FAVGRCKCNGHASECVKN----EFDKL 301

Qy   267 TGVAYTSECFPCKPGTYADKQGSSFCKLCPANSYSNKGETSCHQCDPDKYSEKGSSS--- 323
          || | | | | | | | | | | | | | | | |
Db   302 MCNCKHNTYGVDCEKCLPFFNDRPWRRATAEASASESLPCDCNGRSQECYFDPELYRS--- 358

Qy   324 --------CNVRPA---------CTDKDYFYTHT-----ACDANGETQLMYKNAKPKICSEDLE 365
             | | | | | | | | | | | | |
Db   359 -----TG--HGGHCTNCRDNTDGAK----CERCREN-FFRLGNTEA--CSPCHCSPVGSLSTQC 408

Qy   366 GAVKLPASGVKTHCPCPONPGFFKTNNST---COPCPVGSYSNGSDCTRCPAGTEPAVGFE 422
          | | | | | | | | || | | | | | | | | | | |
Db   409 DSYGRCSCCKPGVMGDKCDRCQPGFHSLTEAGCRPCSCDLRGSTDECNVETGRCVCKDNVE 468

Qy   423 -YKWWNTLPTNMETTVLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILLTLVVPGFRP 481
          | | | | | | | | | | | | | | | |
Db   469 G-------FNCERCKPGFFNLESSNPKGCYPC--FCFGHSSVCTN-----------AVGYS 509

Qy   482 PQSVMADT--ENKEVARITFVFETLCSVNCELYF-----------MVGVNSRTN 522
          | | | | | | | | | |
Db   510 VDISSTFQIDED-------------GWRV--------------------EQRDGSEASLEWSSDRQ 543

Qy   523 TPVETWKGS-KGKQSYTYIIEENTTSFTW-------AFQ------RTTFHEASR 563
          || | | | | | ||| | | | | | |
Db   544 DIAVISDSYFPRYFIAPVKFLGNQVLSYGQNLSFSFRVDRDTRLSAEDLVLEGAGLRVS 603

Qy   564 KYTNDVA---------------KIYSINVTNV 580
           | | | | | |
Db   604 VPLIAQGNSYPSETTVKYIFRLHEATDYPWRPALSPFEFQKLLNNLTSIKIRGTYSERTA 663

Qy   581 MNGVASYCRP-----CALEASDVGSSCCTSCPAGYYIRDSGT----CHSCP------P 623
          | | | | | || | | | || | | | | |
Db   664 GYLDDVTLQSARPGPGVPATWVESCTCPVGYGYGQOFCETCLPGYRRETPSLGPYSPCVLCT 723

Qy   624 NT--ILKAHQPYGV--QACVPCGPGTKNNKIHSLCYNDCTFSRNTPTR------TPNYN 672
          | | | | || | | | | | | | | |
Db   724 CNGHSETCDPETGVCDCRDNTAGPHCEKCSDGYYGDSTLGTSSDCQPCPCGGSSSCAIVP 783

Qy   673 FSALANTVTLAGGPSFTSKGLKYFHH-------FTLSICGNQGRKMSVCTDNVT 719
          || | | | | | | | | | | | || | |
Db   784 KTKEVVCTHCPPTGAGKRCELCDDGYFGDPLGS--------NGPVRLCRPCQCNDNID 833

Qy   720 DLRIPEG---ESCFSKSITAYVCQOAVIIPPEVTGYKAGV------------SSQPVSLADRLI 767
          | | | | | | | | | | | | || | |
Db   834 PNAVGNCNRLTG-----ECLKCIYNTAGFYCDRCKEGFFGNPLAPNPADKCKACACN-- 885

```
Db   886  ----PYGTVQQQSSCNPVTGQ-CQCL--  -PHVSGRDCGTCDPGYYNLQSGQGCERCDCHA  937
           |:|                   ||      ||
Qy   768  GVTTDMTLDGITSPAELFHLESLGIPDVIFFYRSNDVTQSCSSGRSTTIRVRC-------  820

Db   938  LGSTNGQCDIRTGQCE-----  -COPGITGHQGCERCETNHFGF  973
Qy   821  SPQKTVP-------GSLLL----PGTCS---  -DGTCDGCNFHFL---WESAAACPLC  859
Db   974  GPEGCKPCDCDHHEEGSLSLQCKDDGRCECREGFVGNRCDQCQCEENYFYNRSWPGCQCEPAC  1032
```

RESULT 3

MMFB2

laminin gamma-1 chain precursor - fruit fly (Drosophila melanogaster)

N;Alternate names: laminin chain B2
C;Species: Drosophila melanogaster
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000
C;Accession: A31483; A33737; S01733; A40502
R;Chi, H.C.; Hui, C.F.
J. Biol. Chem. 264, 1543-1550, 1989
A;Title: Primary structure of the Drosophila laminin B2 chain and comparison with human,
A;Reference number: A31483; MUID:89109164; PMID:2912972
A;Accession: A31483
A;Molecule type: mRNA
A;Residues: 1-1639 <CHI>
A;Cross-references: EMBL:M25063; NID:g157803; PIDN:AAA28664.1; PID:g157804
R;Montell, D.J.; Goodman, C.S.
J. Cell Biol. 109, 2441-2453, 1989
A;Title: Drosophila laminin: sequence of B2 subunit and expression of all three subunits
A;Reference number: A33737; MUID:90037237; PMID:2808533
A;Accession: A33737
A;Molecule type: mRNA
A;Residues: 1-39,'T',41-891,'L',893-1106,'T',1108-1459,'HV',1462-1581,'G',1583-1639 <MON
A;Note: 831-Tyr was also found
R;Chi, H.C.; Hui, C.F.
Nucleic Acids Res. 16, 7205-7206, 1988
A;Note: the authors translated the codon GGC for residue 409 as Phe
R;Chi, H.C.; Juminaga, D.; Wang, S.Y.; Hui, C.F.
DNA Cell Biol. 10, 451-466, 1991
A;Title: Structure of the Drosophila gene for the laminin B2 chain.
A;Reference number: A40502; MUID:91299161; PMID:1840513
A;Accession: A40502
A;Molecule type: DNA
A;Residues: 1-891,'L',893-1639 <CH3>
A;Cross-references: GB:M58417; NID:g157805; PIDN:AAA28665.1; PID:g157806
C;Genetics:
A;Gene: lamB2
A;Cross-references: FlyBase:FBgn0002528
A;Map position: 3L 67C
A;Introns: 65/3; 110/2; 153/1; 358/1; 495/2; 1357/2; 1469/3; 1570/1
A;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C;Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-297/Product: laminin gamma-1 chain #status predicted <MAT>
F;34-1639/Domain: VI <DOM6>
F;298-528/Domain: V <DOM5>
F;299-356/Domain: laminin-type EGF-like homology <LE01>
F;359-411/Domain: laminin-type EGF-like homology <LE02>
F;414-458/Domain: laminin-type EGF-like homology <LE03>
F;461-511/Domain: laminin-type EGF-like homology <LE04>
F;514-523/Domain: laminin-type EGF-like homology #status atypical <LE05>
F;529-705/Domain: III <DOM3>
F;706-741/Domain: IV <DOM4>
F;710-741/Domain: laminin-type EGF-like homology #status atypical <LE06>
F;744-790/Domain: laminin-type EGF-like homology <LE07>
F;793-844/Domain: laminin-type EGF-like homology <LE08>
F;847-899/Domain: laminin-type EGF-like homology <LE09>
F;902-953/Domain: laminin-type EGF-like homology <LE10>
F;956-1001/Domain: laminin-type EGF-like homology <LE11>
F;1004-1047/Domain: laminin-type EGF-like homology <LE12>
F;1058-1639/Domain: I/I <DOM1>
F;1058-1639/Region: heptad repeats
F;57-67/Disulfide bonds: #status predicted
F;115,147,376,669,862,965,1070,1156,1394,1479,1584/Binding site: carbohydrate (Asn) (
F;1015,1037,1053,1631/Disulfide bonds: interchain #status predicted

Query Match   4.1%; Score 224.5; DB 1; Length 1639;
Best Local Similarity 18.6%; Pred No. 2.3e-06;
Matches 191; Conservative 98; Mismatches 342; Indels 395; Gaps 52;
```

```
Qy    34  TAFQVTQGTGPEL------------HACKESEYHYEYTACDSTGSSRWRVAVPHTPGLCTSL---PD   84
           :  :                    ||
Db   174  TIYKRTSESGPWIPYQFYSATCRDT------YSLPDSRAIRKGEGEAH--ALCTSEYSDIS   226

Qy    85  PVKGTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELD   144
           |||                    |||                                 |
Db   227  PLRDGEIAFS----------TLEGRPS---GINFERSGELQEWTAIDIRITLD   267

Qy   145  LN------------------DSAAESTGNCT--SSKWVP----RGDYIAFNTDECT   174
           ||                  :|:|                  |:|
Db   268  RLNTFGDELFGDSQVLKSYFYAISDIAVGARCKCNGHASKCVPSTGMHGE------   317

Qy   175  ATLMYAVNLKQSGTVNFEYYYPDSSIIFEFFVQNDQCQPNADDSRWMKTTEKGWEFHSVE   234
           :|:                                                      :||
Db   318  RLLVCECRHNTDG-------DRCLPLYNDLKWRST----STE   352

Qy   235  LN----------RGNNVL----------YWRTTAFSV   251
           ||          |||
Db   353  VNECKACNCNGLADKCFFDANLENRTGHGGHCLDCRENRDGPNCERCKENFYMRDDGYCV   412

Qy   252  WTKVPKPVLVRNIAITGVAYTS-EC-----FPCKPGTYADKQGSSFCKLCPANSYSNKGE   305
           ||                    ||      |:|
Db   413  NCACDPVGSRSLLQCNSHGKCQCKPGVTGDK----CDRCD-NNYYQFGP   455

Qy   306  TSCHQCDPDKYSEKGSSSCNVRPACTDKDYFYTHTACDANGETQLMYKWAKPKICSEDLE   365
           ||||                                   ||
Db   456  HGCGQCGCGD-----SGGSHQNT-PAC-----DTETGICF----CKENVE   489

Qy   366  GAVKLPASGVKIHCPPCNPGFF---KTNNSTCQPCPGYSYSNGSDSCTRCPAGTEPAV--   419
           |                                                        |||
Db   490  G-------RCNECKPGFFNLDKNRFGCTPC--FCYGHTSECMTAPGYSIVSVTSN   537

Qy   420  -GFEYKWWNTLPTNMEFTVLSGINFEVKGMTGWEVAG-DHIYTAAGASDNDFMILTLVV   476
            |:|                                  |:|
Db   538  FNKFKERWTAADLNQREVDIKYN---QYSRSIGTTAQGNEHV-   577

Qy   477  PGFRPPQSVMAD---TENKEVARITFVEFTLCSVNCELYFMVGVNSRTNTPVETWKGS-   531
           ||:|                                                     :||
Db   578  -FQAPDRFLGDQRASYNRDLKFKLQLVGQVANTGVSDVILEGAGSRISLPIFA-QGNGI   634

Qy   532  -KGKQSYTYIEENTTTSFTWAFQRTTFHEASRKYTNDVAKI--------YSINVTNVM   581
            |||                                      |||
Db   635  PDQGVKEYTFRLHEH--HDYQWQPS------QSARGFLSTLSNLFAIKIRATYSVQGEALL   687

Qy   582  NGV---------ASYCRPCALEASDVGSSCGTSCPAGYYIDRDSG-----TCH   619
           |||                                                |||
Db   688  DDVELQTAHRGAAGHPATWIEQCTCPEGYLGQCFCESCAPGYRHSPARGGPFMPCIPDCH   747

Qy   620  S-----CPPNT--ILKAHQPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTFTFNYNF   673
           |     :|:|                                                :||
Db   748  GHADICDSETGRCICQHNTHG-DNCDQCAKGFYGNALGG-TPNDC--KRCP------   794

Qy   674  SALANTVTLAGGPSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPEG-----   726
           |||                                                  |||
Db   795  -----CPNDGACLQINEDIVICTECPKGYFGSRCE   824

Qy   727  -ESGFSKSITAYV-----CQAV----IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTL   775
```

```
Db  825 OCSDGGFFGDPGTGLLGEVQTCKRSCDCNGNGNVDPNAVGNGNRTTGECLKCIHNTAGEHGDQCL 884
Qy  776 DG--ITSPAELFHLESLGIPDVIFFYRSNDVTQSCSGGRSTTIRVRCSPQKTYPGSLLLPG 834
Db  885 SGHFGDPPALPH----GRCDRCSCYEAGTEQDEOSIIRCDQVTGQCQCKPNVIGRDC--G 938
Qy  835 TCDGCNFHFLWESA 853
Db  939 ECQPGYFNIRSGNGCENCLCDPVGSYNSTCDRYSGQCHCRPGVMGQRCDQOCENYFYGFSS 998
Qy  854 AACPLC 859
Db  999 EGCKPC 1004
```

RESULT 4
A35502
major surface-labeled trophozoite antigen precursor - Giardia lamblia
C;Species: Giardia lamblia
C;Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 09-Sep-1997
C;Accession: A35502
R;Gillin, F.D.; Hagblom, P.; Harwood, J.; Aley, S.B.; Reiner, D.S.; McCaffery, M.; So, M
A;Title: Isolation and expression of the gene for a major surface protein of Giardia lam
A;Reference number: A35502; MUID:90280395; PMID:2352929
A;Accession: A35502
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-713 <GIL>
A;Cross-references: GB:M33641; NID:g159131; PID:g159132
A;Keywords: surface antigen; transmembrane protein

Query Match   3.8%; Score 211.5; DB 2; Length 713;
Best Local Similarity 21.6%; Pred. No. 6e-06;
Matches 155; Conservative 72; Mismatches 278; Indels 211; Gaps 39;

```
Qy    4 PGHSHHLSARVGRGTERRIPRLWRLLWAGTRAFQVTQGTGPELHACKESEYHYEYTAC-D 62
Db   99 PGHSLCLSSDDGVCTEAAPGYFAPVGAANTEQSV-----IACGD 138

Qy   63 STGSRWRVAVPHTPGLC----TSLPDPVKGTE-----CSFSCNAGEFLDMKDQSCKPCA 112
Db  139 TTGVTIAAGGNTYKGIADCACSAPDATAGAEAGKVATCT-KCGVSKYL--KDNVCVDKA 195

Qy  113 EGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAAESTGNCTSSKWVPRGDYIAPNTDE 172
Db  196 Q---CNSGSTNKFVAVDDSENGNKCVSCSDNLNGGVA----NCDTCSYDEQSKKI-----K 244

Qy  173 CT-ATLMYAVNLKQSGTVNFEYYYPDSSIIFEFFVQNDQCQPN----ADDSRWMKTTEKGW 228
Db  245 CTKCTDNNYLKTTSEGT---------SCVQKDQCKDGFFPKDDS---- 279

Qy  229 EFHSVELNRGNNVLYWRTTAFSVWTKVPKEPVLVRNIAI-----TGVAYTSECFPCKPG 281
Db  280 ------SAGNKCLPCNDSTDG-----IANCATCALVSGRSGAALVT-CSACTDG 321

Qy  282 --TYADK---QGSSFCKLCPANSYSNKGETS--CHQCDPDKYSEKGSSSSCNVRPACTDKD 334
Db  322 YKPSADKTTCEAVSNCKTPGCKACSNEGKENEVCTDCD----GSTYLTPTSQCID-- 372

Qy  335 YFYTHTACDANGETQLMYKWAKPKICSEDLEGAVKLPASGVKT----HCPPCNPGFFKT 389
Db  373 --SCARIGNYYGATEGAK-KLCKE-----CTAANCKTCDDQOGOQACNDGFYK- 417

Qy  390 NNSTCQPCPYG-----SYSNGSDCTRCPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEY 445
Db  418 NGDACSPCHESCKTCSAGTASDCTECPTG----------KALRYGD 453

Qy  446 KGMTGWEVAGDHIYTAAGASDNDFMILTLVVPGF-----RPPQSVMADTENK--EV 494
Db  454 DGTKG--TCGEGCTTGTGTGAGACKTCGLITDGASYCSECATTTEYPQNGVCAPKASRATPT 511

Qy  495 ARITFVFEFLCSVNCELYFMV--GVNSRTNTPVETW------KGSKGKQSYTYIIEENTT 546
Db  512 CNDSPIQNGVCGTCADNYFKMNGGCYETVKVYPGKTVCISAPNGGTCQKAADGYKLDSGTL 571

Qy  547 TSFTWAFQRTTFHEASRKYTNDVAKIYSINVTNVMNGV---ASYCRPCALE--ASDVGS 600
Db  572 ------TVCSEGCKECAS-----STDCTTCLDGYVKSASACTKCDASCETCNGAAT 616

Qy  601 SCTSCPAGYY-IDRDSGTCHSCPPNTILKAHQPYGVQACVPCGPGTKNNKIHSLCY 655
Db  617 TCKACATGYKTASGEGACTSCESD-----SNGVGIKGCLNCAP-PPNNKGSVLCY 667
```

RESULT 5
S34583
serine proteinase (EC 3.4.21.-) PC6B - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C;Accession: S34583
R;Nakagawa, T.; Murakami, K.; Nakayama, K.
A;Title: Identification of an isoform with an extremely large Cys-rich region of PC6,
FEBS Lett. 327, 165-171, 1993
A;Reference number: S34583; MUID:9332934; PMID:8335106
A;Accession: S34583
A;Status: Preliminary
A;Molecule type: mRNA
A;Residues: 1-1548 <NAK>
A;Cross-references: GB:D17563; NID:g407344; PIDN:BAA04507.1; PID:d1005033; PID:g44037
A;Keywords: hydrolase; serine proteinase

Query Match   3.8%; Score 209; DB 2; Length 1548;
Best Local Similarity 19.8%; Pred. No. 2.2e-05;
Matches 231; Conservative 90; Mismatches 399; Indels 446; Gaps 67;

```
Qy   41 GTGPE----------LH-------- 62
Db  320 GPGPDHCSDCLHYYYKLKNNTRICVSSCPPGHYHADKKRC-----RCAPNCESCFGS 372

Qy   82 LPDDPVKGTECSFSCNAGEFLDMKDQSC-KPCAEGRYSLGTGIRFDEWDELPHGFASLSAN 140
Db  373 -HGDQC-LSCKYGYFLNEETSSCVTQCPDGSYE------DIKKNVCGKCSEN 416

Qy  141 MELDDSAAESTGNCTSSKW-------VPRDYIAFNTDECTATLMYAVNLKQSGT-- 188
Db  417 CK----ACIGFHNCTECKGGLSLSLQGSRCSVTCEDGQFFNGHDCQPCHRFCATCSGAGADG 472

Qy  189 -VNFEYYYPDSSIIFEFFVQNDQCQPNADDSRWMKTTEKGWEFHSVELNRGNNVLYWRTT 247
Db  473 CINCTEGY-----VMEEGRCVQSCSVSYLDHSSEG-GYKSCK-RCDNSCLTCNGP 521

Qy  248 AFSVWTKVPKPVLVRNIAITGVAYTSECFPCKPGTYADKQGSSFCKLCPAN------ 298
Db  522 GFKNCSSCPSGYLLD----LGTCQMGAI--CKDGEYIDDQG--HCQTCEASCAKCWGPTQ 573

Qy  299 ------SYSNKGETSCHQCDPDKYSEKGSSSSCNVRPACTDK-- 333
Db  574 EDCISCPVTRVLDDGRCVMNCPSNKFEFRKKQCHPCHYTCQCGQSSGPSNCTSCRADKHGQ 633

Qy  334 DYFYTHTAC-------QLMYKWAKPKICSEDLEGAVKLPASGVK 376
Db  634 ERFLYHGECLENCVGHYPAKGHTCLPCPDNCELCY-----NPHICSRCMSGYVIPPNHT- 689

Qy  377 THCP--PCNPGFFKTNN-STCQPCPYGSYSNGSD---CTRCPAGTEPAVGFEYKWWNTL 429
Db  690 -CQKLECRCGEPQDSEYEECMPCEEGCLGCTEDDPGACTSCATG---YYMFERHCYKAC 744

Qy  430 PTNMETTVLSGINFEYKGM-TGWEVAGDH-------IYTAAGASDNDFMILTLVVPGFR 480
Db  745 P----EKTF--GVKWECRACGTNCGSCDQHECVWCEEGFFLSGGSCVQD-----CGPGFH 793

Qy  481 PPQSVMADTENKEVARITFVFEFLCSVNCELYFMVGVN--SRTNTPVETWKGSKGKGKQSYT 538
Db  794 GDQEL---GECKP------CHRACETCTGSGYNGCVSSCQEGLQLWHG------- 831
```

```
RESULT 6
T26972
hypothetical protein Y47H9C.4 - Caenorhabditis elegans
A;Reference number: Z20293
submitted to the EMBL Data Library, October 1998
R;Harris, B.
C;Accession: T26972
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T26972
A;Molecule type: DNA
A;Residues: 1-1111 <WL>
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Accession: T26972
A;Species: Caenorhabditis elegans
A;Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4
A;Experimental source: clone Y47H9C
C;Genetics:
A;Gene: CESP:Y47H9C.4
A;Map position: 1
A;Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

Query Match 3.6%; Score 197.5; DB 2; Length 1111;
Best Local Similarity 19.9%; Pred. No. 8.2e-05;
Matches 218; Conservative 83; Mismatches 329; Indels 467; Gaps 65;
```

```
Qy  539 YIIEENTTTSFTW-AFQRTFFHEA----------SRKYTNDVAKIYSINVTNVMN 582
Db  832 TCLWSTWPQVEGKDWNEAVPTEKPSLVRSLLQDRRKWKVQIKR----DATSONQ 881

Qy  583 GVASYCRPCALEASDVGSSCTSCPAGYYI----------DRDSGTCHSC------ 621
Db  882 PCHSGCKTCN----GSLCASCPTGMYLWLQACVPSCPQGTWPSVTSGSCEKCSBEDCVS 935

Qy  622 ----------PPNTILKAHQ--------PYGVQA----------CVPG-GPGTKNNK 649
Db  936 CSGADLCQQCLSQPDNTLL-LHEGRCYHSCPEGFYAKDGVCEHCSSPCKTCEGNATSCNS 994

Qy  650 I----HSLCYNDCT----------FSRNTPTRT----------FNYN------ 672
Db  995 CEGDFVLDHGVCVCWKTCPEKHVAVEGVCKHCPERCQDCIHEKTCKECMPDFFLYNDMCHRS 1054

Qy  673 -----FSALANTVTL-----AGGP----------SFTSKGLKYFHHFTLSL----- 703
Db 1055 CPKSFYPDMRQCVPCHKNCLECNGRPKEDDCKVCADYSKAL----HNGLCLDECPEGTYKE 1110

Qy  704 -----CGNQGRKMSVCTDNVTDLRIPEGESGFSKSITAYVCQAVIPPEVTGY----KAG 754
Db 1111 EENDECRDCPESCLICSSAWTCLACREG----FTVVHDVCTA--PKECAAVEYWDEGS 1162

Qy  755 VSSQPV-SLADRLIGVTTDMTLDGITSPAELFHLESLGIPDVIFFYRSNDVTQSC----S 809
Db 1163 HRCQPCHKKCSRCSGPSEDQC--YTCPRETFLLNNTTCVKECPBGYHTDKDSQQCVLCHS 1219

Qy  810 SGRSTTIRVRCSPQKTV----------PGSLLLPGTC----------SDGTCDGCN---- 845
Db 1220 SCRT----CEGPHSMQCLSCRPGWFQLGKECLLQCRDGYYGESTSGRCEKCDKSCKSC 1273

Qy  846 ----------FHFLWESAA-----ACPLCSVADYHA-IVSSCVAGIQKTTYVWREPKL 887
Db 1274 RGPRPTDCQSCDTFFFLLRSKGQCHRACPEHYADQHAQTCERCHPTCDK------ 1323

Qy  888 CSG----------GISLPE 896
Db 1324 CSGKEAWSCLSCVWSYHLLKGICIPE 1349
```

```
Db  272 GALCENECSVGFFGSSGCTGQKDCDLLNNQNCDSSGGECKCIGWTGKHCDIGCSRGRF----G 327
Qy  122 IRFDEWDELPHGFASLSANMELDDSAAE----STG--NCTSSKWVPRGD------YIAF 168

Db  328 LQCKQNCTCP----------GLEFSDSNASCDAKTGCQCCGCSGYKGPCCDERKCDAEQYGAD 379
Qy  169 NTDECTA----TLMYAVNLKQSGTVNFE-YYYPDSSIIFEFFYQNDQCQPNADDSRWMKT 223

Db  380 CSKTCTCVRENTLMCAPN---TGFCRCRKPGFYGDNC---ELACSKDSYGPNCEKQ---- 428
Qy  224 TEKGWEFHSVELNRGNNVLYWRTTAFSWWTKVPKPVLVRNIAITGVAYTSECFP----- 277

Db  429 ----AMCDWN---------HASECNPETGSCV 447
Qy  278 CKPGTYADKQGSSFCKLCPANSYSNKGETSCHQCQPDKYSEKGSSSCNVRPACTDKDYFY 337

Db  448 CKPG----RTGKNCSEPCPLDFY---GPNCAHQC------QCNOR------- 479
Qy  338 THTACD-ANGETQLMYKWAKPKICSEDLEGAVKLPASGVKTHCPPCNPGFFKTNNSTCQ- 395

Db  480 -GVGCDGADGCKCQCQCDRGWTGHR-CEH------------DFFGANCEK 516
Qy  396 --PCPYGSYSNGSD----CTRCPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEYKGM 448

Db  517 RCKCPKGI---GCDPITGECT-CPAG------------LQGANCDIGCP 549
Qy  449 TGWEVAGD--HIYTAAGASDNDFMILTLVVPGFRPPQSYVMADIENKEVARITFVFETLCS 506

Db  550 EGSYGPGCKLHCKCVNGKCDKETCGECT-CQPGFFGSDCSTTCSKGK-------YGESCE 600
Qy  507 VNCELYFMVGVNSRTNTPVETWK----GSKG----KQSYTYIIEENTTTSFTWAF 553

Db  601 LSCPC----SDASCSKQTGKCLCPLGTKGVSCDKCDPNTFGFLCQETVTPS---- 648
Qy  554 QRTFHEASRKYTNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSSCT-SCPAGYYID 612

Db  649 ----------PCASTDPKNGVCLSCPP----GSSGIHCEHNCPAGSYGD 683
Qy  613 RDSGTC----HSCPPNTILKAHQPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTR 667

Db  684 GCQQVCSCADGHGCDPTT----------GECI-CEPG------------YHGKTCSEKCPDG 722
Qy  668 TFNYNFS----ALANTVTLAGGPSFTSKGLKYFHHFTLSLC------GNQGRKMS 712

Db  723 KYGYGCALDCPKCASGSTCDHINGLCICPAGLEG----ALCTRPCSAGFWGNGCRQVC 776
Qy  713 VCTDNVTDLRIPEGESGFSKSITAYVCQAVIIPPEVTGYKAGVSSVSQPVSLADRLIGVTTD 772

Db  777 RCTSEYKQCNAQTGECS----------CPA--------GFQGDRCDKPCE----- 808
Qy  773 MTLDGITSPAELFHLESLGIPDVIFFYRSNDVTQSCS-SGR----------S 813

Db  809 --DGYYGPDCIKHCKCQG----------TATSSCNRVSGACHCHPGFTGEFCHALCPE 854
Qy  814 TTIRVRCSPQKTVPGSLLLPGTCSDGTCDGCNFHFLWESAAAACPLCSVADYHAIVSSCVA 873

Db  855 STFGLKCSKECPKDG----CGDG----YECDAAIGCC----HVDQMSC-- 890
Qy  874 GIQKTTYVWREPKLCSGGISLPEQRVTICKTIDEWLKVGISAGTCTAILLTVLTCYFWKK 933

Db  891 GKAKQEFEALNGAGRSTGLI---WFFVLLIVALCGGLGIAL---FYRN 933
Qy  934 NQKLEYKYSKLVMNATLKDCLPAADSCAIM---EGEDVEDDL------IFTSKN 979

Db  934 --KYQK-----EKDPDMPYSFHKAPNNDEGREFQNPLYSRQSVFPDSDAFSSEN 981
Qy  980 HSLGRSNHL--PPRGLL 994

Db  982 N---GNHQGGPNGLL 994
```

```
RESULT 7
T28811

Qy   39 TQGTGPE--LHACKESEYHYEY----------TACDSTGSRWRVAVPHTPGLCTSLPDPVK 87
Db  224 SDGWGGEFCLNKCEEGKFGAECKFECNCQNGATCDNTNGKC---------ICKS--GYH 271

Qy   88 GTECSFSCNAGEF----------LDMKDQSCK---------PCAEGRYSLGTG 121
```

```
hypothetical protein C54D1.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T28811
R;Minx, M.
A;submitted to the EMBL Data Library, January 1996
A;Reference number: Z20527
A;Accession: T28811
A;Description: The sequence of C. elegans cosmid C54D1.
A;Accession: T28811
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-1557 <MIN>
A;Cross-references: EMBL:U46673; PIDN:AAC48152.1; GSPDB:GN00028; CESP:C54D1.5
A;Map position: X
A;Gene: CESP:C54D1.5
A;Introns: 84/3; 127/1; 166/2; 271/2; 331/1; 392/3; 433/2; 585/2; 1089/1; 1530/3
A;Experimental source: strain Bristol N2; clone C54D1
C;Genetics:
C;Superfamily: laminin beta-1 chain; laminin-type EGF-like homology

Query Match    3.5%;    Score 192;    DB 2;    Length 1557;
Best Local Similarity   19.3%;    Pred. No. 0.00028;
Matches   211;    Conservative   99;    Mismatches 390;    Indels 394;    Gaps   54;

Qy    20  RRIPRLWRLLLWAGTAFQVTQGTGPELHACKESEYHYEYTACDSTGSRWRVAVPHTPGLC   79
Db   114  QQYPTTNLITLVLGKSFDITYVRLKFISPRPESFTIYKKTHTDSEWEPWQF--------  164

Qy    80  TSLPDPVKGTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSA  139
Db   165  --------YSGSCRA------------------------DRAPILPGNEATAQC       191

Qy   140  NMELDDSAAESTGNCTSSKWVPRGDYIAFNTDE----CTATLMYAVNLKQSGTVNFE--  192
Db   192  TKEFSDISPITGGNIAFSTLEGRPSAHAFEESEVLQKWVTASAI-RISLNRMNFGDEVP  250

Qy   193  -----YYYPDSSIIFEFFV------------------QN------                208
Db   251  KDPQVLRSYY----AISDFAVGGRCKCNGHASECVGSSSVDGENRLVCRCEHNTQGADC  306

Qy   209  DQCQPNAADSRWNKTTEKGWEFFHSVELN------                          260
Db   307  NECLPFYNDRPWRSGT------SVEANECIACNCSOLSNRCYFDQOLFEE---------  350

Qy   261  VRNIAITGVAYTSECFPCKPGTYADKQGSSFCKLCPANSYSNKGETSCHQCDPDKYSEKG  320
Db   351  ----TG--HGGHCIDCQGNT----QG-VHCEQCIANHWRPGENYCVACG---CNEIG   394

Qy   321  SSS------CNVRPACTDK----------DYFY----THTACDANGETQLMYKWAKRKI  359
Db   395  SLSTQCDNEGKCQCCKPGVTGRFCDQCLDGFYDFESTNGCKNCQCETSGSLN----NQPRC  449

Qy   360  CSEDLEGAVKLPASGVKTHCPCPCNPGFF---KTNNSTCQPCPYGSYSNGSDCTRCP--  412
Db   450  DSSSGSGSCCKLNVEG---RQCDKCKPGYFDLSTENQFGCTPC--FCFGHSSICNTADGYFA  505

Qy   413  -------AGTEPAVGFEYKWWNTL--PTNMETTVLSGINFEVKGMTGWEVAGD       456
Db   506  MNVSSVFDQDKQKWAG-QNRIGLQDTQWAELDKAVAVSDTDNSPVYF------VAPE   555

Qy   457  HIYTAAGASDNDPMILTLVVPGFRPPQSV-----MADTENKEVARI--------T      498
Db   556  QFLGDQRSSYNQDLVFTLKVAKHVTNQDVKDIIVGADRQELSTSITAQGNPPPTEAQT   615

Qy   499  FVFETLCSVNCELYF-------MVGVNSRPNTPVEFWKGSKGKQSYTYIIEENTTS    548
Db   616  YRFR-----VHADPYFGWPRINELDFIGTLS-NITAIKIRGTYSYKDIGYLSNVNLGTA  669

Qy   549  -------FTW-------AFQRTTFHEASRKYTNDVAKIYSINV                 577
Db   670  GVAPSAANPKQATWIEHCECLPGFVGQFCESCESGFRRET------KFCGPFNHCIKCDC  723

Qy   578  TNVWNGVASYCRPCALEASDVGSSCTSCPAGYYIDRDSGT--CHSCP-PN--TILKAH   630
Db   724  HHHSNSCEAESGSCICHHNTAGDTCERCAARGYYGDALQGTEEDCQKCGPCPNDGPCILHAD  783

Qy   631  -------QPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTRFFNYNFSALANTVT   681
Db   784  GDVICTECPNGYTGRRCDECSDGYFGNPKDGTECVECACACSGNTDP------NSIGNCDK  836

Qy   682  LAG------GPSFTSKGL--------KYFHHFTLSLCGN                     706
Db   837  ITGECKKCIFNTHGFNCENCKPGYWGDALIEPKGNQCQSCGCFAAGTRRPNNDYTLLECNQ  896

Qy   707  VSSQPVSLADRLIGVTTDMTLDGITSPAELFHLESLGIPDVIFFYRSNDVTQSC--EV-TGY--KAG  754
Db   897  QDGQCD-CLPNVIGIQCDQCAHGFYNITSGLGCOECNCDPLGSEGNTCDVNTGQCQCKPG  955

Qy   755  STTIRVRCSPQKTVPGSLLLPGTCSDGTCDGC--NFHFLWESAAACPLCSVADYHAIVSS  812
Db   956  VTGQ---RCDRCADYHFGFSANG-CQPCDCEYIGS------ENQQCDVNSG-         996

Qy   813  STTIRVRCSPQKTVPGSLLLPGTCSDGTCDGC--NFHFLWESAAACPLCSVADYHAIVSS  870
Db   997  ------OCLCKENVEGR--------------RCDCAENRGITQGCLPC--------DD   1028

Qy   871  CVAGIQKTTYVWRE                                              884
Db  1029  CYTLIQSRVNVFRE                                             1042
```

```
RESULT 8
C42125
trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)
N;Alternate names: CRP72
C;Species: Giardia lamblia
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998
C;Accession: C42125
R;Adam, R.D.; Yang, Y.M.; Nash, T.E.
A;Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170
   Mol. Cell. Biol. 12, 1194-1201, 1992
A;Reference number: A42125; MUID:92186850; PMID:1545800
A;Accession: C42125
A;Molecule type: DNA
A;Residues: 1-677 <ADA>
A;Cross-references: GB:M83934; NID:g159123
A;Experimental source: trophozoites
A;Note: sequence extracted from NCBI backbone (NCBIN:88443, NCBIP:88444); this ORF is
C;Keywords: surface antigen

Query Match    3.4%;    Score 189;    DB 2;    Length 677;
Best Local Similarity   17.0%;    Pred. No. 0.00016;
Matches   167;    Conservative   71;    Mismatches 271;    Indels 476;    Gaps   42;

Qy    30  LWAGTAFQVTQGTGPELHACKESEYHYEYTACDSTGSRWR--VAVPHTPG---LCTSL--   82
Db    80  LFMGGCYKTTDGPGSEI--CTRAEGGL-CTECKTANGLFKNPAATPEKGSECILCSDING  136

Qy    83  -PDPVKGTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEW              127
Db   137  ADGYTGVANCAQCTKSDSNKGAAATCTACAQAGYYKDF--QACSKCD----GT       181

Qy   128  DELPHGFASLSANMELDDSAAESTGNCTSSKWVPRGDYIAFNTDECTATLMYAVNLKQSG  187
Db   182  ------------CLTCETSAAQCT-SCPEGKYL-KGDKSCVNNNGCTGN------      216

Qy   188  TVNFEYYPDSSIIFEFFVQNDQCQPNADDSRWMKITEKGWEFHSVELNRGNNVLYWRTT   247
Db   217  ----TYADPESG-------KCLP------CNTIDQACT                      237

Qy   248  AFSWWIKVPKPVLV----RNIAITGVAYTSECFPCKPGTYADKQGSSFCKLCPANSYSNK  303
Db   238  QCEVDSTTKPKCTNCGGQKMVKTAIDGTTTCVDANGCATSNVDGSHFL------ND    288

Qy   304  GETSCHQCDPDKYS---EKGSSSCNVRPACTDKDYFYTHTACDANGETQLMYKWAKPKIC  360
```

```
RESULT 9
T20406
hypothetical protein E01G6.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Accession: T20406
A;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
A;Molecule type: DNA
A;Residues: 1-1391 <WIL>
A;Cross-references: EMBL:Z69717; PIDN:CAA93531.1; GSPDB:GN00028; CESP:E01G6.1
A;Experimental source: clone E01G6
C;Genetics:
A;Gene: CESP:E01G6.1
A;Map position: X
A;Introns: 44/1; 84/3; 175/1; 389/2; 423/1; 501/1; 1358/1

submitted to the EMBL Data Library, February 1996
R;Lloyd, C.
A;Reference number: Z19270

Query Match       3.3%;  Score 183.5;  DB 2;  Length 1391;
Best Local Similarity   18.3%;  Pred. No. 0.00086;
Matches 188;  Conservative 114;  Mismatches 396;  Indels 329;  Gaps 49;
```

```
Db  289 GSTKCILCSDDSESLEANKGTPGCK----------------------------------TCKKNG--------AKP-TC 324
Qy  361 SEDLEGAVKLPASGVKT------------------------------HCPPCNPGFFKTNNSTCQPCPYGSYS 403
Db  325 SECLDGYYNSGNGGTVTCEACGANCANICTQAGNDKCTKCKPGFFMKGNGPTGECVACDNA 384
Qy  404 NGS---------------------DCTRCPAGTEPAVGFEYKWWNNLLPTNMETTVLSGINEYKGM 448
Db  385 QGGIDGCAECTKESTGPLKCTKCKPNRKP----------------- 413
Qy  449 TGWEVAGDHIYTAAGASDNDFMILTLVVPGFRPPQSVMADTENKEVARITFVFETLCSVN 508
Db  414 ----------------AGTSDN-----------------------------YICTEKT 426
Qy  509 CELYFMVGVNSRTNTPVETWKGSKGKQSYTYIIEENTTTSFTWAFQRTTFHEASRKYTND 568
Db  427 CENPTACG--------------GTAGSCDAIVIDDQGTTKHY--------- 454
Qy  569 VAKIYSINVTNVMNGVASYCRP-------CALEASDVGSSCTSCPAGYYIDRDSGTCH 619
Db  455 ----------------CSYCGDSSQAPIDGLCASEAQKAGNTCA------------NGVCT 487
Qy  620 SCPPNTILKAHQPYGVQOACVPCGPGTKNNKIHSLCYNDCTFSRNPTRTNYNFSALANT 679
Db  488 QCTNNYFLYMGGCYSTQK-----APGS------ 509
Qy  680 VTLAGGPSFTSKGLKYFHHFTLSLCGNOGRKMSVCTDNVTD-LRIPEGESGFSKSITAYV 738
Db  510 ----------------FMCKTAGNTG--------ICTEAANNRYFVVPGASNTDQSVLA- 544
Qy  739 CQAVIIPPEVTGYKRAGVSSOPVSLADRLIGVTTDMTLDGITSPAELFHLESLGIPDVIFF 798
Db  545 ----------------CSNPLG--------TLTGTGDTAKAY----VGV------- 565
Qy  799 YRSNDVTQSCSSGRSTTIRVRCSPQKTVPGSL--LLPGTCSDGTCDGC----NFHFLWE 851
Db  566 ----------EGCS--------QCTAPAALSDGGMAPAVCT--SCDSSKKPN----R 598
Qy  852 SAAACPLCSVADYHAIVSSCVAGIQKTTYVWREPKLCSGGISLPEQRVTICKTIDFWLKV 911
Db  599 DGSGCVLCSVGGCKSCVMDNICG-----ECNSGFSLDNGKCVSSGANRSGLSA 646
Qy  912 GISAGTCTAIL------LTVLTCYFW 931
Db  647 GAIAGISVAVAAVVGGLVAFLCWWF 671
```

```
Db  152 YVCSKSQYLNKFICCSNPDG--------EIMGSCPGGETPLENIQCSATKPCNGFSCN- 202
Qy   99 EFLDMKDQSCKP----------CAEGRYSLGTGIRFDEWDELPHGFASLS-ANMELDDSAAESTGN 153
Db  203 ----DKKCCPYASACPAGAPLNNGPMKCCSENAPCPEGHSCVSNAGIQYCCPAKEKVCN 256
Qy  154 CTSSKWPRGDYIAFNTDECTATLMYAVNLKQSGTVNFEYYYPDSSIIFEFFVQNDCQP 213
Db  257 -------MPRNSGV------QC------ASSMKSSN-----RYYFDIGTATCRSF-KFTQCGG 294
Qy  214 NADDSRWMKTTEKGWEFHSVELNRGNNVLYWRTTAFSVWTKVPEKPVVLVRNIAITGVAYTS 273
Db  295 NANN-------------FGSLEECEG----------FCVDTQCPNGQAHRVGAVNANCALA 332
Qy  274 ECPFCPKGTYADKGGSSFCKL-------------CP-----ANSYSNKG--------- 304
Db  333 QADTC------------PNQHRCQOLFLFGPSPICCPTPELTCNEMVSAGTPCCFGRGLTIQRYF 384
Qy  305 ETSCHOQCDPDKY--------SEKGSSSCNVRPACTDKDYFYHTACDA-------NGETQLMYKW 354
Db  385 NPSTQKCCQPFHYGCSGNGNNFETV-DOCQNFCLHSADSACDGLVPLKNPNSELOR----- 439
Qy  355 AKPKICSEDLEGAVKLPASGVKTHCPCNPGFEFKTNNSTCQPCPYGSYSNGSDCTR---C 411
Db  440 -----CSEE-----------DPCPAGYECNDSSYC-CPSSENACNANMSRGNGC 476
Qy  412 PAGTEPAVGFEYKWWNNTLPTNMETTVLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMI 471
Db  477 KGSTQRSMWFYDK--------------SKKKCSQFVYNGCGGTPNRFTT 511
Qy  472 LTLVVPGFRP-----------PQSVMADTENKEVARITFVFE----------- 502
Db  512 KVACTESCVQSSAAFGLCPRGMNPFTEEGEITPKTCTLNVRSTCPSGASCVKSSTNQSICC 571
Qy  503 --TLCSVNCELYFMVGVNSRINTPVETWKGSKGKQSYTYIIEENTTTSFTWAFQRTT-- 557
Db  572 KAVTACPDNRIPVNIPGTNSVVACSIE--------RDDCPDDFQCLESAFAPG 616
Qy  558 FH-------EASRKYTNDVAKIYSI------NVTNVMNGVASYCRPCALEASDVGSSC 602
Db  617 FHMCCSGPLPSAQPVRPIAPVAPMPRRPFEPDSFADLLSNIS----PCPPQLFSSGQSC 672
Qy  603 T-------SCPAGYYIDRDSG-----TCHSCPPNTILKAHQPY-----GVQAC------- 638
Db  673 TVNQVGDCPKNHICFRDIGFKQGSCCRTTPPKCSQRGYVPVFLDRTQVQCQVDLDGCPE 732
Qy  639 -----------VPC----GPGTKNNKIHSLCYNDCTFSRNPTRTFNYNFSALANTVT 681
Db  733 DAKCMTSSVAKLAICCQKYQPPSPNNNRPANRGVNQGSVSNNNQN--NNKGAVCANGET 789
Qy  682 LAGGPSFTSKGLKYFHHFTLS--LCGNQGRKMSVCTDNVTDLRIPEGESGFSKSITAYVC 739
Db  790 PFRGPDNKFQECNFVHNTCPSGYQCEFSSTGQAVCCNTNVIRCPSGSSVFEYGGRPLAC 849
Qy  740 QAVIIP-PEVTGYKAGVSSQP-----VSLADRLIGVTTDMTLDGI------TSPAELFHLE 788
Db  850 PAGSNKCPQ--GYSCMPSTNPQHHLCCSSGSGMGVSQPCLRGVAYVNPATNQRQFCSPM 907
Qy  789 SLGIPDVIFFYRSNDVTQ--SCSSG-RSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDGCN 845
Db  908 KADCPAGYTCFESDQSSQFICCTHGDLSDRFQGYCPPRQ-----IPYISRDGFPPTCH 960
Qy  846 FHFLWESAAACPLCSVADYHAIVSS------------CVAGIQKTTYVWREPKLCSGGISLPEQRV 899
Db  961 MQL-----SPCP--TTAPYVCIYSPEKQDSYCCAPIDTAVHVVNPDR---GYASPAKNI 1009
Qy  900 TICKTID 906
Db 1010 DLLETPD 1016
```

```
RESULT 10
T21889
```

```
Qy   47 HACKESEYHYETACDSTGSRWRAVAVPHTPGLCTSLPDPVKGTECS------FSCNAG 98
```

hypothetical protein F36H2.3b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R;Steward, C.
C;Accession: T21889
submitted to the EMBL Data Library, October 1996
A;Reference number: 219483
A;Accession: T21889
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1797 <WIL>
A;Cross-references: EMBL:Z81078; PIDN:CAB03077.1; GSPDB:GN00019; CESP:F36H2.3b
A;Experimental source: clone F36H2
C;Genetics:
A;Gene: CESP:F36H2.3b
A;Map position: 1
A;Introns: 33/1; 150/1; 257/1; 358/3; 416/3; 479/1; 510/1; 581/3; 652/3; 722/3; 793/3; 8

Query Match   3.3%;   Score 179.5;   DB 2;   Length 1797;
Best Local Similarity   19.5%;   Pred. No. 0.0021;
Matches 206;   Conservative 98;   Mismatches 367;   Indels 384;   Gaps 53;

```
Qy   42 TGPELHACKESEYHYEYTACDSTGS----------------------RWRVAVPHTP---------GL  78
         |||                                            |||              :|
Db  147 TGP-----------HYAYTFNDVGQSLPGYEKEFFPLCFAWR----QTPSVVLFNSTDDGV 192

Qy   79 CTSLPDPVKGTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFEDEWDELPHGFASLS 138
         | |||                                                     |
Db  193 CLTLPE------------VQNGKILYSNDQL-------NVFSIGTSATL----ECHQGFA--- 229

Qy  139 ANMELDDSAAESTGN-------CTSSKWVPRGDYIAFNTDECTATLMYAVNLKQSGTVNFE 192
                 ||        |  |:|                 |:|  |: |
Db  230 ----------GNGPSSLVCTKNGWVPKRE-------DLGSCVRLEQSNK----- 261

Qy  193 YYYPDSSIIFEFFVQNDQCQPNADDSRWMKTTEKGWEFHSVELNRGNNVLY-----WRTT 247
            ||                                                  :||
Db  262 ---SRLLVASD-VTPTSSSCRVPSSTPNG--------------------NIVYSANVATSTS 299

Qy  248 AFSWWTKVPKPVLVRNIAITGVAYTSECFPCKPGTYADKQGSSFCKL---CPANSYSNKG 304
         |                            |                    |:|
Db  300 INSVPTATRATVLCSLGHVPTTSVTSS--KCVDGEWEPSLPTCLSLLDIKCPILSAPRNG 357

Qy  305 ETSCHQCDPDKYSEKG---SSSC-----------NVRPACTDKDYFYTHTACDANGETQLMYKW 354
         |                    | :|                    :||
Db  358 ELVFTNSVKSPYSLNSVISLKCDRNYFGTGNLTSTCTSTGWDQKIGQCEPVGIRRLSESA 417

Qy  355 AKPKICSEDLEGAVKLPASGVKTHCPPCNPGF-FKTNNSTCQPCPYGSYSNGSDCTRCPA 413
                  ||        |  :||:         |
Db  418 STGAVCA-----AIANPANGNLLYM-QSNPTVQYSSGTSAYLMCNLGYSLSGGSVSTLCSN 471

Qy  414 GT-EPAVGFEYKWWNTLPTINMETTVLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMIL 472
          |                  |  |
Db  472 GVWSPSIG----QCTNALALGQTTGNCEAI-------PTRINGTITYSSPG--------- 511

Qy  473 TLIVPGFRPEQSVWADTENKEVARITFVEETLCSVNCELYFMVGVNSRTNTPVEIWKGSK 532
                  |  |                :|
Db  512 ----------TYISGTTIATLACNLMNTVSGSSTSTCLSGVWNPTI 546

Qy  533 GKQSTYIIEENTTFSFTWAFQRTTFHEASRKYTNDVAKIYSINVTNVMNGVASY----- 587
          |||                                                   |:|
Db  547 G-----------------NCVSSGGTGGTGTTCPNPT--------VINQOITYNQGNT 579

Qy  588 --CRPCALEASDVGSSCTSCPAGYYIDRDS-----------GTCH--------SC 621
                                              |  ||        ||
Db  580 FDIRPSGTAAT-------LTCNSGYTISGSSQSTCTNGAFSPTLGTCNFGSSTGGAGTTC 633

Qy  622 PPNTILK---------AHQPYGVQOACVPCGPG-TKNKIHSLCYN--------DCTF 660
          |   :                                               |
Db  634 PNPTIVNGQVTYNQGNTFDDATRPALTIATLTCNSGYTISGTSTCTNGVGFTPTLGTCTF 693

Qy  661 SRNTPTPRTFNYNFSALANTVTLAGG------PSFTSKGLKYFHHFTLSLCGNQGRKMSV 713
                                         |  |                  |
Db  694 GTSGSGTTACINPIIMNGQITYSQGNTYDITRPSGTVATLTCNSGYTIS----GSSQST 748
```

RESULT 11
T21888
hypothetical protein F36H2.3a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R;Steward, C.
C;Accession: T21888
submitted to the EMBL Data Library, October 1996
A;Reference number: 219483
A;Accession: T21888
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1805 <WIL>
A;Cross-references: EMBL:Z81078; PIDN:CAB03076.1; GSPDB:GN00019; CESP:F36H2.3a
A;Experimental source: clone F36H2
C;Genetics:
A;Gene: CESP:F36H2.3a
A;Map position: 1
A;Introns: 33/1; 150/1; 257/1; 358/3; 416/3; 479/1; 510/1; 581/3; 652/3; 722/3; 793/3

Query Match   3.3%;   Score 179.5;   DB 2;   Length 1805;
Best Local Similarity   19.5%;   Pred. No. 0.0021;
Matches 206;   Conservative 98;   Mismatches 367;   Indels 384;   Gaps 53;

```
Qy  714 CTDNVTDLRIPEGESGFSKSITAYVC-QAVIPPEVTGYKAG----VSSOPVSLADRLI-- 767
         |                :                                 |
Db  749 CTNGAFSPTLGTCNFGGSTGGTGTTCPNPTVVNGQIT-YSQGNTFDATRPALTTATLTCN 807

Qy  768 ------GVTTDMTLDGITSPAELFHLESLG----------IPDVIFFYRSN 802
               :                 |                         ||
Db  808 SGYTISGTSISACMNGVFTP----TLGTCTFGSSGSSAGTACMNPYVINGQITYSQGN 860

Qy  803 -DVTQSCGSSGRSTTIRVRCSPQKTVPGS-------LLLP-GTCSDGTCDG----CN 845
              |                         |      |||:  |
Db  861 TVDITRPA-----GTTATLTCSSGYTVSGDYSVSTCTNGIFPTPTLGTCTLGSSTGGTGIQC 915

Qy  846 FHFLWESAAACPLCSVADYH-----------------------AIVSSC----VAGIQKTTYWREPKLC 888
                                                      |||
Db  916 -----TAMIAPLGGSVIYSNGGTMGPPSGTTVTGSCTNGGAITGASTAT-------C 961

Qy  889 SGGISLPEQRVTICKTIDFWLKVGISAGTCTAIIL 923
          :|                             |
Db  962 SNGMWNPIFLGT-CSLIG----GSTTCQCSALTI 990

Qy   42 TGPELHACKESEYHYEYTACDSTGS----------------------RWRVAVPHTP---------GL 78
         |||                                            |||              :|
Db  147 TGP-----------HYAYTFNDVGQSLPGYEKEFFPLCFAWR----QTPSVVLFNSTDDGV 192

Qy   79 CTSLPDPVKGTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFEDEWDELPHGFASLS 138
         | |||                                                     |
Db  193 CLTLPE------------VQNGKILYSNDQL-------NVFSIGTSATL----ECHQGFA--- 229

Qy  139 ANMELDDSAAESTGN-------CTSSKWVPRGDYIAFNTDECTATLMYAVNLKQSGTVNFE 192
                 ||        |  |:|                 |:|  |: |
Db  230 ----------GNGPSSLVCTKNGWVPKRE-------DLGSCVRLEQSNK----- 261

Qy  193 YYYPDSSIIFEFFVQNDQCQPNADDSRWMKTTEKGWEFHSVELNRGNNVLY-----WRTT 247
            ||                                                  :||
Db  262 ---SRLLVASD-VTPTSSSCRVPSSTPNG--------------------NIVYSANVATSTS 299

Qy  248 AFSWWTKVPKPVLVRNIAITGVAYTSECFPCKPGTYADKQGSSFCKL---CPANSYSNKG 304
         |                            |                    |:|
Db  300 INSVPTATRATVLCSLGHVPTTSVTSS--KCVDGEWEPSLPTCLSLLDIKCPILSAPRNG 357

Qy  305 ETSCHQCDPDKYSEKG---SSSC-----------NVRPACTDKDYFYTHTACDANGETQLMYKW 354
         |                    | :|                    :||
Db  358 ELVFTNSVKSPYSLNSVISLKCDRNYFGTGNLTSTCTSTGWDQKIGQCEPVGIRRLSESA 417

Qy  355 AKPKICSEDLEGAVKLPASGVKTHCPPCNPGF-FKTNNSTCQPCPYGSYSNGSDCTRCPA 413
                  ||        |  :||:         |
Db  418 STGAVCA-----AIANPANGNLLYM-QSNPTVQYSSGTSAYLMCNLGYSLSGGSVSTLCSN 471
```

```
Qy   414  GT-EPAVGEFEYKWWNTLPINMETTVLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMIL 472
               |:||
Db   472  GWWSPSIG--QCTNALALGQTTGNCEAI----------PTRINGTITYYSPG------- 511

Qy   473  TLVVPGFRPQSVMADTENKEVARITFVEFTLCSVNCELYFMVGVNSRTNTPVETWKGSK 532

Db   512  ----------TYTSGTIATLACNLMNTVSGSSTSTCLSGVWNPTI 546

Qy   533  GKQSYTYIIEENTTTSFTWAFQRTTFHEASRKYTNDVAKIYSINVTNVMVGVASY----- 587

Db   547  G---------NCVSSGGTGGTGTTCPNPT----------VINQITYNQGNT 579

Qy   588  --CRPCALEASDVGSSCTSCPAGYYIDRDS----------GTCH----------SC 621

Db   580  FDITRPSGTRAAT------LTCNSGYTISGSSQSTCTNGAFSPTLGTCNFGSSTGGAGTTC 633

Qy   622  PPNTILK---------AHQPYGVQACVPCGPG-TKNNKIHSLCYN-----DCTF 660

Db   634  PNPIVVNGQVTYNQGNTFDATRPALTIALTCNSGYTISGTSTSTCINGVFTPLGTCTF 693

Qy   661  SRNPTTRTFENYNFSALANIVTLAGG-----PSFTSKGLKYFHHFTLSLCGNQRKMSV 713

Db   694  GTSGGTGTACINPIIMNGQITYSQGNTYDILTTRPSGTVATLTCNSGYTIS----GSSQST 748

Qy   714  CTDNVTDLRIPEGESGFSKSITAYVC-QAVIIPPEVTGIKAG-----VSSQPVSLADRLI 767

Db   749  CTNGAFSPTLGTCNFGSSTGGTGTTCPNPTVVNCQIT-YSQGNTFDATRPALTTATLTCN 807

Qy   768  -----GVTTDMTLDGITSPAELFHLESLG---------IPDVIFFYRSN 802

Db   808  SGYTISGTSISACMNGVFTP------TLGTCTFGSSGSAGTACMNPVYINGQITYSQGN 860

Qy   803  -DVTQSCSSGGRSTTIRVRCSPQKTVPGS------LLLP--GTCSDGTCDG-----CN 845

Db   861  TYDITRPA------GTTATLCSSGYTVSGISVSTCTNGIFTPTLGTCTLGSSTGGTGIOC- 915

Qy   846  FHFLWESAAACPLCSVADYH------------AIVSSC-----VAGIQKTTYVWREPKLC 888

Db   916  -----TAMIAPLGGSVISNGGTMGPFPSGTTVTVGSCTNGGAITGASTAT--------C 961

Qy   889  SGGISLPEQRVTICKTIDFWLKVGISAGTCTAILL 923

Db   962  SNGMWNPTFLGT-CSLIG-----GSTTGQCSALTI 990
```

RESULT 12
T42017
cysteine rich protein - Giardia intestinalis
C;Species: Giardia intestinalis
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: T42017
R;Chen, N.; Upcroft, P.; Upcroft, J.
Parasitology 111, 423-431, 1995
A;Title: A Giardia duodenalis gene encoding a protein with multiple repeats of a toxin h
A;Reference number: Z22027
A;Accession: T42017
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <CHE>
A;Cross-references: EMBL:L29079; NID:g951190; PID:g951191; PIDN:AAA74587.1
A;Experimental source: specific host: Homo sapiens
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

Query Match 3.2%; Score 178; DB 2; Length 1274;
Best Local Similarity 19.9%; Pred. No. 0.0018;
Matches 202; Conservative 93; Mismatches 397; Indels 324; Gaps 51;

```
Qy   79   CTSLPDP--VKGTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEWD-ELPHGF 134

Db   395  CTKDPEAPCNVEGCETCVEGNA------QQCKTCRPG-YTINDTKQCTKDPEAP--- 442

Qy   135  ASLSANMELDDSAAESTGNCTSSKWVPRGDYIAFNTDECTATLMYAVNLKQSGTV--NF 191
```

```
Db   443  -------CNVEGCETCVE--GNAQQCKTCRPGYTINTDTKQCTKDPEAPCNVEGCETCVEGNA 496

Qy   192  EYYYPDSSIIEFFVQNDQCQP------NADDSRWMKTTEKGWEFHSVELNRGNNVLYWRT- 246

Db   497  Q--------QCKTCRPGYTINTDTKQCTKDPEAPCNVEGCETCVEGNAQQCKTC 542

Qy   247  ------TAFSWTKVPK-PVLVRNIAITGVAYTSECPPCKHPGTYADKQGSSPCKLCPAN 298

Db   543  RPGYTINTDTKQCTKDPEAPCNVEGCETCVEGNAQQCKTCRPG-YTINTDTKQCTKDPEA 601

Qy   299  SYSNKGETSC------HQC------DPDKYSEKGSSSCNVRPACTDKDYFYTHTAC 342

Db   602  PCNVEGCETCVEGNAQQCKTCRPGYTINTDTKQCTKDPEAPCNVE-GC------ETCV 652

Qy   343  DANGETQLMYKWAKP-------KICSEDLEGAVKLP-----ASGVKTHCPCNPGFFKT 389

Db   653  EGNAQ---QCKTCRPGYTINTDTKQCTKDPEAPCNVEGCETCVEGNAQQCKTCRP 708

Qy   390  NNST------------CQPCPYGS-----YSNGSPDCTRCPAGTEPAVGFEYK-YTI 424

Db   709  NRDTKQCTKDPEAPCNVEGCETCVEGNAQQCKTCRPGYTINTDTKQCTKDPE- 760

Qy   425  WWNTLPTNME--TTVLSG------INFEYKGMTGWEVAGDHIYTAAGASDNDF 469

Db   761  -------APCNVEGCETCVEGNAQQCKTCRPGYTINTDTKQCTKDPEAPCNVEGCE 816

Qy   470  MLTLVVPGFRPQSVMADTENKEVARITFVEFTLCSV-NCELYFMVGVNSRTNTPVETW 528

Db   817  QQCKTCRPGY-------TINTDTKQCTKDPEAPCNVEGCE- 853

Qy   529  KGSKGKQSYT----YIIEENTTTSFTWAFQRT--------THEASRKYTNDVAK 571

Db   854  -GNAQQCKTCRPGYTI--NTDTK----QCTKDPEAPCNVEGCETCVEGNAQQCKTCRP 904

Qy   572  IYSINVTNVMNGVASYCRPCALEASDV----GSSCTSCPAGYYIDRDSGTCHSCPNTI 626

Db   905  GYTIN-TDTKQCTKDPEAPCNVEGCETCVEGNAQQCKTCRPGYTINTDTKQCTKDP--- 959

Qy   627  LKAHQPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTRTFENYNFSALANTVTLAGGP 686

Db   960  -------EAPCNTPNCKTCDNPKTDNEICTKC-NDGDYL--TPTN------QCVPDCTAISGYY 1007

Qy   687  SFTSKGLKYFHHFTLSLCGNQGRKMSVCTD------NVTDLRIPE------ 725

Db   1008 GDTDKKCKACNPECAECVGPANNQCTACPVGKMLQYTDTNTPVNGGTCMDQCSVSSTNDG 1067

Qy   726  -GESGFSKSITAYVCQAVIIPPEVTGYKAGVSSQPVS--LADRLIGVTTDMTLDGITRSP 781

Db   1068 CAECGAQIGGTAYCSKC--------KNTQQAPLGNCAASSRVAPCATITSGACTKC 1116

Qy   782  AELFLHLESLGIPDVIFFYRS----NDVTQSCSSGGRSTTIRVRCSPQKTV 826

Db   1117 NEGYFLKDGGC------YQTDROPGKQVCSNAQGGNGKCQTCANGLA------- 1157

Qy   827  PGSLLLPGTCSDGTCDGCNFHFLWESAAACPLCSVADYHAIVSSCVAGIQK-----TTYV 881

Db   1158 ------ASDGNCAECH------STCATCSTADAADKCKTCATGYYKENGDDTTA- 1199

Qy   882  WREPKLCSGGISLPEQRVT-----ICKTIDFWL-----KVGISAGTCTAILLTVL 926

Db   1200 -GLCKKCSEKISGCKQCVSSSSSSVICLESEVGTGGSVNKSGLSTGAIAGISVAVI 1254
```

RESULT 13
A42125
trophozoite cysteine-rich surface antigen 170 - Giardia lamblia
N;Alternate names: CRP170; cysteine-rich surface antigen CRP170
C;Species: Giardia lamblia
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 01-Dec-2000
C;Accession: A42125; B42125; S00530; S48056
R;Adam, R.D.; Yang, Y.M.; Nash, T.E.
Mol. Cell. Biol. 12, 1194-1201, 1992

```
A;Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170 ge
A;Reference number: A42125; MUID:92186850; PMID:1545800
A;Accession: A42125
A;Molecule type: DNA
A;Residues: 1-98 <ADA1>
A;Cross-references: GB:M83937; NID:g159124
A;Experimental source: trophozoite
A;Note: sequence extracted from NCBI backbone (NCBIN:88421, NCBIP:88427); this ORF is no
A;Accession: B42125
A;Molecule type: DNA
A;Residues: 1269-1766 <ADA2>
A;Note: sequence extracted from NCBI backbone (NCBIN:88424, NCBIP:88431); this ORF is no
R;Adam, R.D.; Aggarwal, A.; Lal, A.A.; de la Cruz, V.F.; McCutchan, T.; Nash, T.3.
J. Exp. Med. 167, 109-118, 1988
A;Title: Antigenic variation of a cysteine-rich protein in Giardia lamblia.
A;Reference number: S00530; MUID:88089405; PMID:3335828
A;Accession: S00530
A;Molecule type: DNA
A;Residues: 1-56 <YAN>
A;Cross-references: EMBL:L25059
R;Yang, Y.; Adam, R.D.
Nucleic Acids Res. 22, 2102-2108, 1994
A;Title: Allele-specific expression of a variant-specific surface protein (VSP) of Giard
A;Reference number: S48056; MUID:94301794; PMID:8029018
A;Accession: S48056
A;Molecule type: DNA
A;Residues: 1154-1409, 'A', 1411-1420, 'K', 1422-1425, 'R', 1427-1481 <ADA3>
A;Cross-references: EMBL:X06174.; NID:g9355; PID:g929603
A;Note: the source is designated as Giardia intestinalis
A;Experimental source: trophozoites WBA6
A;Comment: This translation was produced by PIR staff from information provided by the a
C;Genetics:
A;Gene: VSPA6
C;Keywords: surface antigen; tandem repeat

Query Match     3.2%;     Score 177.5;     DB 2;     Length 1766;
Best Local Similarity   19.5%;     Pred. No. 0.0028;
Matches 208;     Conservative 102;     Mismatches 357;     Indels 401;     Gaps 60;

Qy    59 TACDSTGSRWRVAVPHTPGLLCTSLPDP-----VKGTECSFSCNAGEFLD--MKD-QSCKPC 111
          |:|   |||                         |: | || |:|  |      ||| :||
Db   880 TKCDANG------AAPYLKK--TNPSDPTGTCVSAVDCQGS--AGYYTDDSVSDAKECKKC 930

Qy   112 -AEGRYSLGTGIRFDEWDELPHGFASL----------------SANMELDDSAAE 149
          |                                          ||  : ||:
Db   931 NAPCTACAGTADKCTKCD--ANGAAPYLKKTNPSDPTGTCVSAVDCQGSAGYYTDDSVSD 988

Qy   150 --------STGNCTSSKWVPRGDYI-AFNTDECTATLMYAVNLKQSGTVNFE 192
           |   || |                 |:| | : || : :: |
Db   989 AKECKKCNAPCTACAGTADKCTKCDANGAAPYLKKTNPSDPTGTCVSAVDCQGSAG----- 1044

Qy   193 YYYPDSSIIFEFFVVQNDQCQP-NADDSRWMKTTEKGWEFHSVELNRGNNVLYWRTTAFSV 251
           |||                   ||                        :
Db  1045 -YYTDDSV-----SDAKECKKCNAPCTACAGTADK------CTKCDANGAAPYLKKTNPS- 1092

Qy   252 WTKVPKPVLVRNIAITGVA--YT-------SECFPCK-----PGTYADKQGSSFCKLCP 296
                                 |           |||       |||
Db  1093 ---DPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKCNAPCTACAGT-ADK----CTKCD 1143

Qy   297 ANS---YSNKGETSCHQCPD----------------KYSEKGSSSCNVRPA 329
          | |         ||                      ||
Db  1144 ANGAAPYLKKTNPS---DPTGTCVSAVDCQGSAGYYTDDSVSDAKECKKCNAPCTACAG 1199

Qy   330 CTDKDYFYTHTACDANGETOLMYKWAKPK----ICSEDLEGAVKLPASGVKTHCPPCN 383
           |||             ||                |  ||  |
Db  1200 TADK-----CTKCDANGAAPYLKK-TNPSDPTGTCVSAVDCQGSAGYYTDDSVSDAKEC 1252

Qy   384 PGFFKTNNSTCQPCPGYGSYSNGSDCTRCPA-GTEPAVGFEYKWWNTLPTNMETTVLSGIN 442
                |||                   |:|    ||
Db  1253 ----KKCNAPCTACA------GTADKCTKCDANGAAPYLK------KTNPSDPTGTCVSAV- 1297
```

```
RESULT 14
A48434
variant-specific surface protein - Giardia lamblia (strain GS/M)
C;Species: Giardia lamblia
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 07-May-1999
C;Accession: A48434
R;Nash, T.E.; Mowatt, M.R.
Mol. Biochem. Parasitol. 51, 219-228, 1992
A;Title: Characterization of a Giardia lamblia variant-specific surface protein (VSP)
A;Reference number: A48434; MUID:92244292; PMID:1574080
A;Accession: A48434
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-557 <NAS>
A;Cross-references: GB:M80480; NID:g159142; PID:g159143

Query Match     3.2%;     Score 177;     DB 2;     Length 557;
Best Local Similarity   20.2%;     Pred. No. 0.00076;
Matches 145;     Conservative 67;     Mismatches 235;     Indels 272;     Gaps 39;

Qy   443 FEYKGMTGW-------------------EVAGDHIYTAAGAS----DNDFMLITLVV 476
          |: | |:                    |   | | |         | |  |:
Db  1298 -DCQGSAGYYTDDSVSDAKECKKCAEGQKPNTAGTQCFSCSDANCERCDQNDVCARC--- 1353

Qy   477 PGFRPPQSVMADTENKEVARITFVFETLCSVNCELYFMVGVNSRINTPVETWKGSKGK-- 534
          ||                          || :  |                   |||:
Db  1354 ------STGAPPENGKCPAATPGCHSSCDGCTE-----NAMTNQADKCTGCKEGRYL 1399

Qy   535 ------QSYTYIIEENTITSFTWAFQRTTFHEASRKYTNDVAKTIYSINVTNVMNGVASY 587
                 |:|   |                             |
Db  1400 KPESAAGQSGTCLTAEECTSD----TTHFTKEKAGD--SKGMCLSCSDATHGITG- 1448

Qy   588 CRPCAL------EASDVGSSCT----SCPAGYYIDRDSG--TCHSCPPNT 625
          | |||       ||:  || |     ||||||:  | |    ||:||  |
Db  1449 CKKCALKTLSGEAESTVVCGSECTDKRLTPSGNACLDNCPAGTYADNINGVSVSCASCHA-T 1507

Qy   626 ILKAHQPYGVQOACVPCGPGTKNNKIHSL-------CYNDCT--FSRNTPTRTFNYNFS 674
           |  | | || :|: | |                   | ||    | | || | :: |
Db  1508 CAECNGDANAASCTACYPG------YSLLYGSGTPAGTCVRECTGAFITNCADGQCTAN- 1559

Qy   675 ALANTVTLAGGPSFTSK-------GLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPE 725
           |                       ||            | |    ||
Db  1560 ------VGGAKYCAQCKDGYAPIDGI-----CTTVAAAGRDASVCT-------A 1595

Qy   726 GESGFSKSITAYVCQOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAELF 785
           | |  |   |  |                  |     |
Db  1596 ADGKCTKCAGEYTL---------MSGGCYGVAKLPGKSVCTL------ 1628

Qy   786 HLESLGIPDVIFFYRSNDVTQSCCSGRSTTIRVRCSPQKTVPGSLLILPGTCSDGTCDGCN 845
                |   :      |:       |                       ||
Db  1629 ------ASNGKCTMCAANNQAPVEERKC-PE--------CSEG-CAKCN 1660

Qy   846 FHFLWESAAACPLCSVADYHAI---------VSSCVA------GIQKT 878
                      | |                  ||| |       || |
Db  1661 -DSNACTECLPGYKGAGDKCFKCTASSGNNNQITGVANCVTCAPPAGNGGPVT 1714

Qy   879 TYWREPKLCSGGITSLPEQRVTICKTIDPWLKVGISAGTCTAILLTVL 926
           :   |  | || :|             :    ||||| ||
Db  1715 CYIKTDGDNTGGSVN-------------KSGLSTGAIAGISVAVV 1746
```

```
Qy   269 VAYTSECFPCKPGTYADKQGSSFCKLCPANSYSNKGETSCHQCPDKYSEKGSSSCNVRP 328
          | |  || | |                   | ||||| ||| |
Db    32 VGETEICTQCKQ-NYVPING----VCEAAASSN--TKCQKASADEASDQ----TCG 76

Qy   329 ACTDKDYFYTHTACDANGETOLMYKWAKPKICSEDLEGAVKLPASGVKTHCPPCN-PGF 386
           |  | :             |               |    ||        |   || |
Db    77 KCLSTTFMYKGGCYDKTGNLGRI----ICKTEGSDAGK-------CGACNEKGF 120

Qy   387 FKTNNSTCQPCPGYGSYSNGSDCTRCPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEYK 446
          | ||   |  |       |  ||  |                          | ||
Db   121 FDNPDAA-------NNVDSCISC--GDATGV-----------------TIPGSSTKT 153
```

```
RESULT 15
T30201

Notch homolog protein - sea squirt (Halocynthia roretzi)
C;Species: Halocynthia roretzi
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: T30201
R;Hori, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.
Dev. Genes Evol. 207, 371-380, 1997
A;Title: Notch homologue from Halocynthia roretzi is preferentially expressed in the cen
A;Reference number: Z20775
A;Accession: T30201
A;Molecule type: mRNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-2352 <HOR>
A;Cross-references: EMBL:AB001327; NID:d1204472; PID:d1026501; PIDN:BAA25571.1
C;Genetics:
A;Gene: Notch

Query Match      3.2%;  Score 175;  DB 2;  Length 2352;
Best Local Similarity  18.4%;  Pred. No. 0.0057;
Matches  204;  Conservative 128;  Mismatches 381;  Indels 398;  Gaps 64;


Qy  447 GMTGWEVAGDHIYTAAGASDNDFMILTLVVPGFRPQSVMADTENKEVARITFVFETLCS 506
            ::|                                              :||
Db  154 GVAGCAKC---------------------------------TKPSQISENTGTKEA---- 180
                                                      |:|  ||   |    TCTE 180

Qy  507 VNCELYFMVGVNSRTNTPVETWKGSKGKQSYTYIIEENTTFSFTWAFQRTTFHEASRKYT 566
        |:   |                                  |:  ||
Db  181 CNANLY-LKAVSSPTS---------------------ATSCVSAEDCKTIGY---FPTTDTTDSKKK-- 221

Qy  567 NDVAKIYSINVTNVMNGVASYCRPCAL-------------EASDVGSSCTS------C 605
                              | |                  : ||  |
Db  222 -------CLTCSTADKGGIDGCSACELLPSTTRASTVLISCSACSTNNLSPLKNECMQDC 274

Qy  606 PAGYIDRD--------SGTCHSCP-PNFILKAHQPYGVQACVPCGP----GTKNNKIHSL 653
        ||  |            |   |                      ||
Db  275 PAGTYADSNVCKPCHTSCASCKGDNT----ESSCTACYPGSVLSYGTDNTK-GT 323

Qy  654 CYNDCTFSRNTPTRTFNYNFSALANTVTLAGGPSFTSGLKYFHHFTLSLCGNQGRKMSV 713
        |                                                          |
Db  324 CIABCT------------------------GKYLEN 335

Qy  714 CTDNVTDLRIPEGESGFSKSITAYVCQAVIIPPEVTGY-----KAGVSSQPVSLADRLIGV 769
        |               |                          |    |
Db  336 CADGQCTATI-AGSKYCSK-----CKSGFVP--VNGLCVSAETARAAPPGSTPDKTNGV 386

Qy  770 TTDMILDGITSPAELFHLESLGIPDVIFFYRSNDVTQSCSGSRSTTIRVRCS--PQKTVP 827
        |                                                          |
Db  387 CTACT-------------EKYFLESGGCYQAEKF-PGNTLCTTADAGKCTT---CANGQDKDSN 433

Qy  828 GSL-LLPGTCSD---------GTCDGCNHFLWESAAAACPLCSVADYHAI-VSSCVAGIQKTT 879
        |    |                 |                                |
Db  434 GSCPACPTNCASCAKDNTKTCNKCFSGYYLTDTAKACKKCSETISGNIQGVENCISCLAPTS 493

Qy  880 YVWREPKLC------------SGGISLPEQRVTICKTIDFWLKVGISAGTCTAILLTV 925
                              |                                 |
Db  494 PQSSTPVTCYVKTSGGGSGGSGDNSTGGDSGPN-----LSSGAIAGISVAVIVV 540


Db  252 VYGF-----------------TRDDCSENIDDCSNVAC-----FHNARCIDQAGTFECLCTPGNR 294
        |:                                                            |
Qy  241 VL------------------YWRTTAFSVWTIKVPKPVLVRNIAITGVAYTSEC 275

Db  295 ILCHLDDACISDPCARGATACDTTNPITGHHMCDCPDGWTDKDCSRDIDECSLGG------ 347
        | |                                             |
Qy  276 FPCKPGTYADKQGSSFCKLCPANSYSNKGETSCHQCDPDKYSEK------GSSSCNVRP 328

Db  348 NPCEHNGQCNNTDGSFECICVAGYSGPRCETNINECEPNPCRNDATCLDMIGNFNCVCMP 407
                                                                     |
Qy  329 ACT--------DKDYFYTHTACDANGETQL-----MYKWA----KPKICSEDLEGAVKLPASGV 375

Db  408 GFTGIICDEDIDECESNPCANGGTCIDEVNAYTCSCALGFTGDDCSQNIDECASTPCMNK 467
        |   |                                                      |
Qy  376 KT---------HCPPCNPGF-----FKTNNSTC--QPCPYGS------------401

Db  468 ATCIDKANAYEC-ECAPGYTGVHCETNIDDCVINPCHYGSCRDGVNTFYCDCLLGYEGTK 526
            |                                                      |
Qy  402 --------YSNGSDCT----------RCPAGT-----EPAVGFEYKWWNTLPTN 432

Db  527 CQIDTNECASSPCENGGTGTDEIGYVTCTCPGTGSGSSCEINPDDCVGNPCQY 579
        |                                                   |
Qy  433 METTVLSGINFEYKGMTGWEVAGDHIYTAAGASD------NDFMILTLVVPG 478

Db  580 -GTCVDGVD-DYSCSCTPGYTGEHCDTDIINECDSNPCMNGATCQNEVNNFVC----- 629
            |                                                   |
Qy  479 FRPQSVMADTENKEVARITFVFETLCSVNCELYFMVGVNSRTNTPVETWKGSGKGKQSYT 538

Db  630 -QCPFGIMGTQCSSDIQE----CSSNPCLHEYARRDQHVHCICDA-GYQGENCET 678
            |                                                  |
Qy  539 YIIE--------ENTTTSFTWAFQRTTFHEASRKYTNDVAKIYSINVTNVMNGVASY 587

Db  679 EINECASNPCQHGACENKVAQFV----SHCDAGYTGTACEI-DIN------ 718
        |                                                  |
Qy  588 CRPCALEASDVGSSCCTS--------CPA----GYYIDRDSGTC--HSCPPNTILKAHQPYG 634

Db  719 -ECATQPCQNGGTCTSGINSYNCACPAKYTGVNCETELSPCVNPCENGATCQESADYL 776
            |                                                   |
Qy  635 VQACVPCGPGTK--------NNKIHSLCYN-----DCTFSRNTPTRTFNYNFSA 675

Db  777 AYVC-QCPEGFRGPPCATDIINECVNSPCKNGGGCTNLVPGYQCTCSQGFGTGKDCDTDID 835
        |  |                                                       |
Qy  676 LANTVTLAGGPSFTSGLKYFHHFTLSLCGNQGRKMSVCTDNVTD-LRIPEGESGF-SKS 733

Db  836 CSSNPCLNGGQCLLDDVG-----SYKCLCLPGFEGNN--CQEEVNECASFPCKNGGICTDY 888
        |                                                          |
Qy  734 ITAYVCQAVIIPPEVTGYKA------GVSSQPVSLADRLIGVT 770

Db  889 VNSYVCTC-----LSGFYSLDCEKNIEDCSSSSCMNGGTVCDGTINSYSCSCTANFTG-- 940
        |                                                          |
Qy  771 TDMILDGITSPAELFHLESLGIPDVIFFYRSNDVTQSCSGSRSTTIRVRC-SPQKTVPGS 829

Db  941 -DKCQNAVNNCASL-QCQNGG----TCYPDSGDPKCACVHGYTGI--HCESLQNLCTG- 990
            |                                                    |
Qy  830 LLLPGTCSDGTCDGCNHFLWESAAAACPLCSVADYHAIVSSCVAGIQKTTYVWREPKL-C 888

Db  991 --PNICKNG------GSC-----VQTSNTVSCNCLGGYEGTDCA--VPQVSC 1027
            |                                              |
Qy  889 SGGISLPEQRVT-ICKTIDFWLKVGISAGTC 918

Db 1028 TVGASLLGIAVSDLC-----LNGGTC 1048


Qy   53 EXHYEYTACDS--TGSRWRVAVPHTPGLCTSLPDPVKGTEC-----SF--SCNAGEF--- 100
        |                                                        |
Db   91 DFSYQCT-CQTGFTGDTCSQVLVCSPNPCSN------GAGCEELSNSFKCTCTSGYYGDT 143

Qy  101 LDMKD--QSCKPCA--EGRYSLGTGTIRFDEWDELPHGFASLSANMELDDSA-- 147
        |                                                   |
Db  144 CANDVNECDTPDICQNACTCSNNDGYSC-----SCVAGFEGNNCEVNIDDCSGH 193

Qy  148 AESTGNC-TSSKVVPRGDYIAFNTDECTATLMYAV--NLKQSG--TVNF 191
        |                                             |
Db  194 SCQNGATCADAVSTYDCHCPAEWT--GQYCTIDVDECSLSNNVAKRRDLQQTEGGFTCNC 251

Qy  192 EVYYPDSSIIFEFFVQNDQCQPNADDSRWMKITEKGWEFHSVE-----LNRGNN 240
        :|                                              ||
```